43

fteval JOURNAL for Research and Technology Policy Evaluation

ISSUE 43 | AUGUST 2017

CONFERENCE »open evaluation 2016 – INTERNATIONAL RTI POLICY EVALUATION CONFERENCE«

openevaluation 2016

INTERNATIONAL RTI POLICY EVALUATION CONFERENCE

FEDERAL MINISTRY FOR SCIENCE, RESEARCH AND ECONOMY

Minoritenplatz 5, 1014 Vienna
Mag.ª Irene Danler
E: irene.danler@bmwfw.gv.at
Mag.ª Simone Mesner
E: simone.mesner@bmwfw.gv.at

Stubenring 1, A -1014 Wien
Mag.ª Sabine Pohoryles-Drexel
E: sabine.pohoryles-drexel@bmwfw.gv.at

FEDERAL MINISTRY OF TRANSPORT, INNOVATION AND TECHNOLOGY

Radetzkystraße 2, 1030 Vienna
Dr. Rupert Pichler
E: rupert.pichler@bmvit.gv.at
Dr. Mario Steyer
E: mario.steyer@bmvit.gv.at

AQ AUSTRIA

Renngasse 5, 1010 Vienna
Dr.in Elisabeth Froschauer-Neuhauser
E: elisabeth.froschauer-neuhauser@aq.ac.at
Dr.in Eva Maria Freiberger
E: eva.maria.freiberger@aq.ac.at

AIT - AUSTRIAN INSTITUTE OF TECHNOLOGY

Tech Gate Vienna,
Donau-City-Straße 1, 1220 Vienna
Mag. Michael Dinges
E: michael.dinges@ait.ac.at
Mag.ª Barbara Heller-Schuh
E: barbara.heller-schuh@ait.ac.at

AWS - AUSTRIA WIRTSCHAFTSSERVICE GESELLSCHAFT MBH

Walcherstraße 11A, 1020 Vienna
Dr. Joachim Seipelt
E: j.seipelt@awsg.at
Mag. Norbert Knoll
E: n.knoll@awsg.at

CDG - CHRISTIAN DOPPLER RESEARCH ASSOCIATION

Boltzmanngasse 20, 1090 Vienna
DIin Brigitte Müller
E: mueller@cdg.ac.at

CONVELOP COOPERATIVE KNOWLEDGE DESIGN GMBH

Bürgergasse 8-10/I, 8010 Graz
Mag. Markus Gruber
E: markus.gruber@convelop.at

FFG - AUSTRIAN RESEARCH PROMOTION AGENCY

Sensengasse 1, 1090 Vienna
DIin Dr.in Sabine Mayer
E: sabine.mayer@ffg.at
Mag. Leonhard Jörg
E: leonhard.joerg@ffg.at

FWF - AUSTRIAN SCIENCE FUND

Sensengasse 1, 1090 Vienna
Dr. Falk Reckling
E: falk.reckling@fwf.ac.at
Dr.in Dorothea Sturn
E: dorothea.sturn@fwf.ac.at

IHS - INSTITUTE FOR ADVANCED STUDIES

Josefstädter Straße 39, 1080 Vienna
Mag. Richard Sellner
E: richard.sellner@ihs.ac.at

JOANNEUM RESEARCH FORSCHUNGSGESELLSCHAFT MBH

Sensengasse 1, 1090 Vienna
Mag. Wolfgang Polt
E: wolfgang.polt@joanneum.at
Mag. Jürgen Streicher
E: juergen.streicher@joanneum.at

AUSTRIAN INSTITUTE FOR SME RESEARCH

Gusshausstrasse 8, 1040 Vienna
Dr.in Sonja Sheikh
E: s.sheikh@kmuforschung.ac.at

LUDWIG BOLTZMANN GESELLSCHAFT

Nußdorfer Str. 64, 1090 Vienna
Dr. Peter Mayrhofer
E: peter.mayrhofer@lbg.ac.at

AUSTRIAN COUNCIL FOR RESEARCH AND TECHNOLOGY DEVELOPMENT

Pestalozzigasse 4/DG 1,
1010 Vienna
Dr. Johannes Gadner
E: j.gadner@rat-fte.at
Dr.in Constanze Stockhammer
E: c.stockhammer@rat-fte.at

ÖAW - AUSTRIAN ACADEMY OF SCIENCE

Dr. Ignaz Seipel-Platz 2, 1010 Vienna
Nikolaus Göth, MSc
E: nikolaus.goeth@oeaw.ac.at

TECHNOLPOLIS AUSTRIA

Rudolfsplatz 12/11, 1010 Vienna
Mag.ª Katharina Warta
E: warta@technopolis-group.com

WIFO - AUSTRIAN INSTITUTE OF ECONOMIC RESEARCH

Arsenal, Objekt 20, PF 91, 1103 Vienna
Dr. Jürgen Janger
E: Juergen.Janger@wifo.ac.at
Dr. Andreas Reinstaller
E: andreas.reinstaller@wifo.ac.at

WWTF - VIENNA SCIENCE AND TECHNOLOGY FUND

Schlickgasse 3/12, 1090 Vienna
Dr. Michael Stampfer
E: michael.stampfer@wwtf.at
Dr. Michael Strassnig
E: michael.strassnig@wwtf.at

WIRTSCHAFTSAGENTUR WIEN. EIN FONDS DER STADT WIEN

Mariahilferstraße 20, 1070 Vienna
Robert Mayer-Unterholzer
E: mayer-unterholzner@wirtschaftsagentur.at

ZSI - CENTRE FOR SOCIAL INNOVATION

Linke Wienzeile 246, 1150 Vienna
Dr. Klaus Schuch
E: schuch@zsi.at
MMag. Alexander Degelsegger
E: degelsegger@zsi.at

TABLE OF CONTENT

ISSUE 43 | 2017

EVALUATION OF MISSION ORIENTED RESEARCH

INNOVATION TAXONOMY, DESIGNS AND ASSESSMENT APPROACHES 2

INDIVIDUALS AND CAREER TRAJECTORIES

CONNCEPTS AND APPROACHES TOWARDS OPENNESS

NEW EVALUATION APPROACHES, TOOLS AND SOURCES

EVALUATING SMART SPECIALISATION AND REGIONAL INNOVATION

STAKEHOLDER INVOLVEMENT IN RESEARCH PRODUCTION AND CITIZEN SCIENCE EVALUATION

EVALUATING IMPACTS OF SSH

EVALUATION OF INNOVATION REGULATIONS, PROGRAMMES AND INSTRUMENTS

POSTER SESSION

EDITORIAL

KLAUS SCHUCH, MANAGING DIRECTOR OF FTEVAL AND STEFAN PHILIPP, ASSISTANT TO THE MANAGEMENT OF FTEVAL

DEAR READERS,

This issue of the fteval Journal for Research and Technology Policy Evaluation contains the proceedings of the Open Evaluation conference, organised by the Platform and its partners from the Manchester Institute of Innovation Research (MIoIR) and the Institute for Research and Innovation in Society (IFRIS) in November 2016. You find a collection of the extended abstracts presented at the conference. In an upcoming issue you will also find a selection of full papers.

This editorial also gives us the possibility to reflect the impressions, things learned and challenges identified at the conference.

As regards our focus on 'open evaluation', we think that we have directed attention to a concept which is about to emerge, but which still lacks sufficient empirical testing and evidence. The topic "openness" in STI policy evaluation will doubtlessly accompany the STI policy evaluation community in the next couple of years. We noticed good will and pretence but at the same time also lack of ambiguity related to practices and approaches. We assume that more empirical evidence on the notion of openness in evaluation will be available in a few years. It is expected that this will go hand in hand with the use of open data and novel data science approaches and the pursuit of identifying impact along different impact dimensions, including societal impact of STI policies. However, also more risk-taking by funding agencies, especially public ones, is required to experiment with alternative open approaches in controlled settings. This might cause additional costs but also can bring additional benefit and might transcend the action spaces of conventional STI policy evaluations.

We will for sure follow-up on this and believe that our next conference on measuring impact of R&D and its many bifurcations will provide a good interface and point of contact.

Finally, I can confess that we feel lucky that we managed to organise the largest conference in Europe dedicated to the evaluation of policies in the field of research, technology and innovation policy (RTI) despite our extremely limited capacities. In comparison to previous conferences we gathered more people, from all over the world, and managed to mobilise more young professionals, also thanks to our cooperation with EU-SPRI. We received an overwhelmingly positive response, which of course has mostly to do with the impressive quality of the speaker's contributions! We are thankful to the many helping hands from the STI policy evaluation community from all over Europe: our key note speakers, the organisers of the panels, the panel discussants, the session chairs, the moderators and the paper presenters and our sponsors, without whom the conference could not have been realised: The Austrian Federal Ministry of Transport, Innovation, and Technology (BMVIT), The Austrian Federal Ministry of Science, Research and Economy (BMWFW), The Vienna Science and Technology Fund (WWTF), The Austrian Science Fund (FWF), The Austrian Research Promotion Agency (FFG), Danube-INCO.NET, project funded under FP7. As mentioned above, our next conference will take place in Vienna under the auspices of the Austrian EU Council Presidency in early November 2018. The focus of this conference will be on impact of R&D, which became the dominant narrative in research and innovation policy-making in Europe.

We hope to see you again! In the meanwhile enjoy reading!

Klaus Schuch

Stefan Philipp

BEYOND THE REF (RESEARCH EXCELLENCE FRAMEWORK)? WHAT DOES THE EVIDENCE TELL US ABOUT DESIGNING A FUTURE PERFORMANCE-BASED RESEARCH FUNDING SYSTEM FOR THE UK AND OTHER COUNTRIES?[1]

ERIK ARNOLD, PAUL SIMMONDS, KRISTINE FARLA, PETER KOLARZ, BEA MAHIEU AND KALLE NIELSEN

The UK's Research Excellence Framework (REF) is the current version of the world's longest-running performance-based research funding system (PRFS). Originally introduced in 1986 as the Research Selectivity Exercise and later Renamed the Research Assessment Exercise, it rapidly became the mechanism used to allocate almost all institutional funding for research to the UK universities. It is conducted every six years or so and is completely based on peer review, though in recent iterations it has to some degree been informed by bibliometric indicators. The last exercise in 2014 cost in excess of €300m to run (Farla & Simmonds, 2015) and although it is widely criticised and even referred to as a 'Frankenstein monster' (Martin, 2011) the UK university community is reluctant to change it (Technopolis, 2010) (Wilsdon, 2015). While, therefore, the REF is an outlier in terms of age, the proportion of institutional funding it steers as well as the focus on peer review (Mahieu & Arnold, 2015) and its cost, a comparison of its apparent impacts and its methods with evidence from other countries will shed useful light not only on the opportunities to improve the UK system but also other systems elsewhere.

Our paper therefore assembles existing published evidence about the REF and other national systems in an attempt to understand their desirable and undesirable effects, the set of design parameters available to PRFS designers and to develop some evidence-based principles of use in improving future designs in the UK and elsewhere. It builds on research we are conducting for the UK Department of Business, Innovation and Skills (to be published in July 2016) as well as earlier work for the STOA office of the European Parliament (Mahieu, Arnold, & Kolarz, 2013) and for the Czech Ministry of Education and Culture (Mahieu & Arnold, 2015), which entailed proposing a new design for the Czech PRFS.

Governments increasingly use PRFS to (i) stimulate efficiency in research activity; (ii) allocate resources based on merit; (iii) reduce information asymmetry between supply and demand for new knowledge; (iv) inform research policies and institutional strategies; and (v) demonstrate that investment in research is effective and delivers public benefits (Abramo, D'Angelo, & di Costa, 2011). In the tradition of New Public Management (Ferlie, Ashburner, Fitzgerald, & Pettigrew, 1996) (Boston, Martin, Pallot, & Walsh, 1996), PRFS seek to increase accountability for the expenditure of taxpayers' money. They are seen as a means for selectively distributing research funds. But most also seek to use them to drive particular behaviours. The shift to performance based funding – whether through PRFS or performance contracts – is part of a broader movement to make universities more autonomous and introduce more strategic university research management.

There is a small but growing body of evidence that confirms that PRFS can have a positive effect on national research performance. A recent JRC report indicates that, while the presence of a PRFS alone does not explain overall performance, the introduction of such systems is generally followed by performance improvement (Jonkers & Zacharewicz, 2016) and the proliferation of such systems since about the year 2000 suggests that policymakers believe this. There are, of course, cases where the steering signals provided by a PRFS lead to undesired effects, such as the famous example of a former Australian system, which encouraged researchers to publish more papers – as a result of which the overall output did indeed rise, while the average quality (measured in bibliometric terms) went down (Butler, 2003). However, at the policy level, the overall view of PRFS is generally positive – even if the differences in policy contexts and policy purposes among countries are barely discussed in either the 'grey' or 'white' literature (Arnold & Barker, 2015).

The scientific literature, however, is generally rather critical of the effects of PRFS on research and researchers.

- Whether peer review or indicator-based, PRFS appear to have a bias against interdisciplinary research, which reflects the inherent biases of these assessment mechanisms (OECD, 2010) (Elsevier, 2015)

1 Parts of our full paper will be based on work currently still in progress, which is due to be reported in July 2016. This abstract therefore relies on work by ourselves or others that is already in the public domain. No part of this abstract should be construed as indicating the conclusions of work so far unpublished.

- An inhibiting effect on radical or 'transformative' research (Laudel & Gläser, 2014)
- The exclusion of 'heterodox' approaches, as in economics (Lee, 2007)
- An increased separation between teaching and research functions (Mahieu & Arnold, 2015)
- A tendency to disadvantage applied research compared with 'basic' research (OECD, 2010)

These effects are reinforced by the influence of university and research management over recruitment and career progression (Aagaard, et al., 2014) (Abramo, D'Angelo, & di Costa, 2011) (Bence & Oppenheim, 2004).

Internationally, there are three overlapping 'waves' of PRFS. The first – started by the UK system – was heavily based on peer review, expensive and burdensome to operate. A second has begun to address the cost and burden problems through greater reliance on academic output indicators. We see the start of a third wave where PRFS begin to incorporate indicators of societal impacts, going beyond the academic sphere. The REF is the leading example here.

Internationally, however, the range of options for implementing PRFS is much wider than that entertained in the UK. This 'design space' is summarised in Table 1.

Table 1: Key design parameters for the assessment component in PRFS

Key design parameter	Variations
Model used for the assessment of research quality	• Peer review-based • Informed peer review • Mix peer review & bibliometrics • Metrics-based
Scope of research activity included	• Research • Innovation • Societal relevance
Type of indicators	• Output indicators • External funding indicators • Systemic indicators • Outcome/impact indicators
Assessment criteria in peer review-based systems	• Quality of outputs • Relevance of research activities • Institutional environment • Esteem measures
Granularity	• Units of analysis (grouping of scientific disciplines) • Inclusion of individual staff (inclusive/exclusive)
Periodicity	• Annual • Longer time frames

Source: derived from (Mahieu & Arnold, 2015)

In our current work, we have defined five scenarios, intended to 'test the envelope' defined by this design space and are discussing these with some 50 UK stakeholders (universities, research funders and other interested parties) in order to establish the degree to which the UK research community would have confidence in alternatives to the current REF design, which might reduce the considerable burden in cost and time on the community while preserving positive aspects of the REF. In parallel, we are assembling evidence from the literature about effects associated with specific design features. For example, wholly indicator-based systems induce different kinds of 'gaming' compared with peer review based systems; the proportion of institutional funding allocated by a PRFS strongly influences the degree to which the assessment technique needs to be beyond criticism (Mahieu & Arnold, 2015).

In our paper, we will be in a position to report

- More comprehensively on the broad body of evidence about the effects of PRFS in general and the REF in particular
- How design alternatives measure up against existing evidence and experience
- The degree to which the UK community, in the context of the status of the REF more or less as an institutionalised part of the context, is prepared to countenance change
- Based on these analyses, a set of trans-nationally applicable design principles for future PRFS

BIBLIOGRAPHY

Aagaard, K., Bloch, C., Schneider, J. W., Henriksen, D., Kjeldager, T. R., & Lauritdsen, P. S. (2014). Evaluering af den Norske Publiceringsindikator,. Aarhus: Dansk Center for Forskningsanalyse.

Abramo, G., D'Angelo, C. D., & di Costa, F. (2011). National research assessment exercises: a comparison of peer review and bibliometric rankings. Scientometrics(89), 929-941.

Arnold, E., & Barker, K. E. (2015). Getting what you measure? Effects of performance-based research funding systems . Atlanta Conference on Science and Innovation Policy. Atlanta.

Bence, V., & Oppenheim, C. (2004). The influence of peer review on the research assessment exercise. Journal of Information Science, 30 (4), 347-368.

Boston, J., Martin, J., Pallot, J., & Walsh, P. (1996). Public Management: The New Zealand Model. Oxfourd: Oxford University Press.

Butler, L. (2003). Explaining Australia's increased share of ISI publications – the effects of a funding formula based on publication counts. Research Policy, 32(1), 143-155.

Elsevier. (2015). A review of the UK's interdisciplinary research using a citation-based approach. A report by Elsevier to the UK HE funding bodies and MRC.

Farla, K., & Simmonds, P. (2015). REF 2014 Accountability Review: Costs, benefits and burden. Bristol: HEFCE.

Ferlie, E., Ashburner, L., Fitzgerald, L., & Pettigrew, A. M. (1996). The New Public Management in Action. Oxford: Oxford University Press.
Hicks, D. (2011). Performance based university research funding systems'. Research Policy, 41(4), pp. 251-261.

Hicks, D. (2012). Performance-based university research funding systems. (Elsevier, Ed.) Research Policy, 41, 251-261.

Jonkers, K., & Zacharewicz, T. (2016). Research Performance Based Funding Systems: a Comparative Assessment, EUR 27837 EN; doi:10.2791/659483. European Union, JRC.

Laudel, G., & Gläser, J. (2014). Beyond breakthrough research: Epistemic properties of research and their consequences for research funding . Research Policy, 43(7), 1204-1216.

Lee, F. (2007). The Research Assessment Exercise, the state and the dominance of mainstream economics in British universities. Cambridge Journal of Economics(31), 309-325.

Mahieu, B., & Arnold, E. (2015). R&D Evaluation methodology & Funding principles - The R&D Evaluation Mthodology (Final Report 1). Technopolis Group. Prague: Ministry of Education, Youth & Sports.

Mahieu, B., Arnold, E., & Kolarz, P. (2013). Measuring scientific performance for improved policy making - Literature review. Technopolis Group. Brussels abd Strasbourg: STOA, European Parliament,.

Martin, B. (2011). The Research Excellence Framework and the 'impact agenda': are we creating a Frankenstein monster? Research Evaluation, 20(3), 247-254.

Martin, B., & Whitley, R. (2010). The UK Research Assessment Exercise: a Case of Regulatory Capture? In R. Whitley, J. Gläser, & L. Engwall, Reconfiguring knowledge production: changing authority relationships in the sciences and their consequences for intellectual innovation (pp. 51-80). Oxford: Oxford University Press.

OECD. (2010). Performance-based Funding for Research in tertiary Education Institutions: Workshop Proceedings. Paris: OECD.

Technopolis. (2010). REF Research Impact Pilot Exercise Lessons-Learned Project: Feedback on Pilot Submissions . Bristol: HEFCE.

Wilsdon, J. (2015). The Metric Tide: Report of the independent Review of the Role of Metrics in Reseaech Assessment and Management. Bristol: HEFCE.

AUTHORS

ERIK ARNOLD
PAUL SIMMONDS
KRISTINE FARLA
PETER KOLARZ
BEA MAHIEU
KALLE NIELSEN

Technopolis Ltd
3 Pavilion Buildings
Brighton BN1 1EE
UK

CORRESPONDING AUTHOR: ERIK ARNOLD
erik.arnold@technopolis-group.com

ENCOURAGING EVIDENCE-BASED STI POLICYMAKING IN JAPAN: OVERVIEWING SCIENCE FOR RE-DESIGNING SCIENCE, TECHNOLOGY AND INNOVATION POLICY (SCIREX)

SHINANO HAYASHI, TATEO ARIMOTO, MASAHIRO KURODA, KEIKO MATSUO AND HIROAKI HARADA

Demand of science for science, technology and innovation (STI) policy is significant in response to fast growing social and economic problems in multi-disciplinary areas. Many countries have been searching ways for designing more effective and efficient STI policies to induce innovation. Since the former science advisor to the President of the United States, Dr. John H. Marburger III, addressed the need to establish the Science of Science Policy in 2005, the Science of Science and Innovation Policy (SciSIP) program of the National Science Foundation has been started for bridging research and practice in this interdisciplinary field. On the other hand, the use of scientific evidence in policymaking has been discussed amongst stakeholders in science policy community as a challenge we face today. Not only delivering practical scientific evidence into policymaking is difficult, but also policy-biased evidence tends to be mingled. Like many other countries, rigorous analyses to present evidence for public decision-making are also needed in Japan. In 2011, Center for Research and Development Strategy of Japan Science and Technology Agency (JST-CRDS) published a strategic proposal which suggested strong needs for promoting co-evolution of science of STI Policy and STI policy system (JST-CRDS, 2011) for enhancing evidence-based policy in Japan.

Encouraged by the proposal, in April 2011, Ministry of Education, Culture, Sports, Science and Technology of Japan (MEXT), launched "SciREX: Science for RE-designing Science, Technology and Innovation Policy" program based upon the designing philosophies such as:

1. Formulate policy based on scientific rationality,
2. Attain scientifically rational policymaking processes,
3. Assure public accountability raising policymaking transparency,
4. Publicly disseminate research results and findings so as to further public participation in policymaking processes,
5. Promote collaboration; clarify division of responsibilities among policymaking participants and contributors.

Steering board of the program regards that developing Science of STI Policy is important as well as improving the science, technology and innovation (STI) policymaking system. Thus, advances in Science of STI Policy support improvements to the system, which in turn support advances in Science of STI Policy. Both are necessary for the process to work in a virtuous circle. As such, co-evolution of Science of STI Policy and STI Policy system is major target of SciREX program, hence the program give emphasis on policy design using analytical results (SciREX Symposium, 2015).

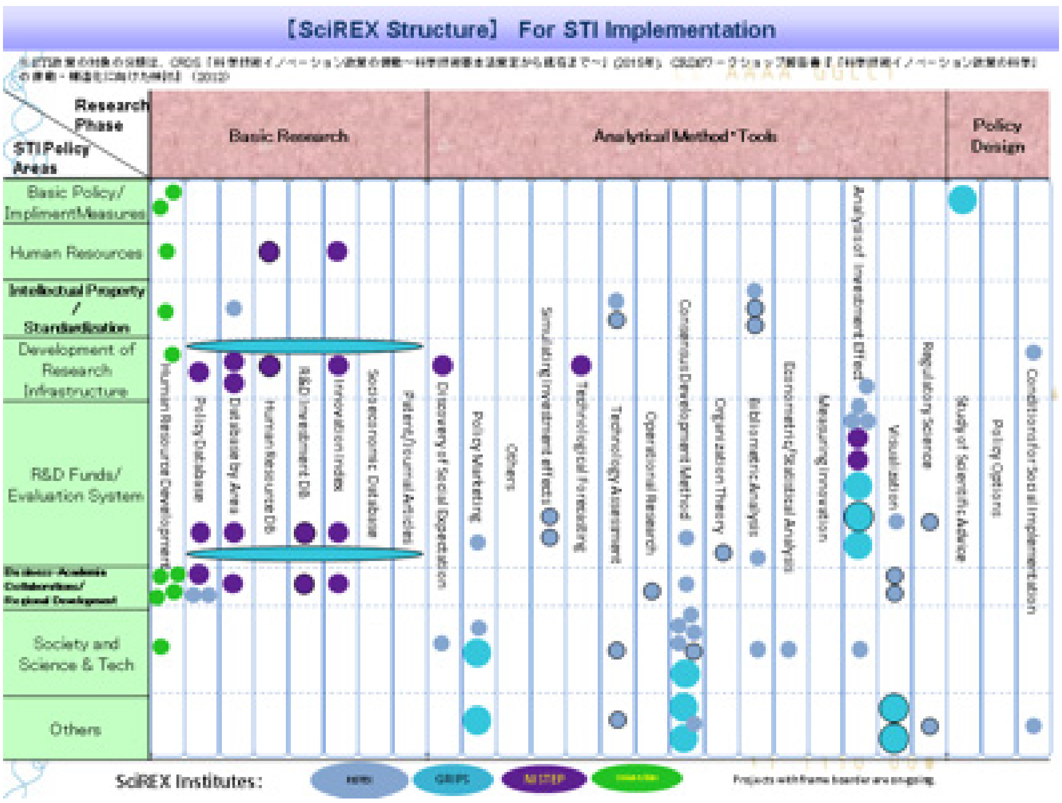
Likewise U.S. National Science Foundation's Science of Science and Innovation Policy (SciSIP) program, which was "seeking to build a community of practice both among researchers in many disciplines in the area of science and innovation policy and between this community and its counterpart (federal) government communities" (Teich and Feller, 2009), SciREX program also has focused upon efficient policymaking process amongst stakeholders thorough science evidence from stringent interdisciplinary research outcomes. Nonetheless, SciREX program unlike SciSIP, not only provides competitive research fund for interdisciplinary studies, but also establishes database for interdisciplinary researches and human resource development programs in several universities. By conducting these activities, it is believed to produce the innovation-inducing interaction of stakeholders. The ultimate goal of this program is to realize "evidence-based policy formation", which tries to make policy more effective in order to address policy challenges, based on observations and analysis of social and economic states from various aspects as well as setting plausible policy options.

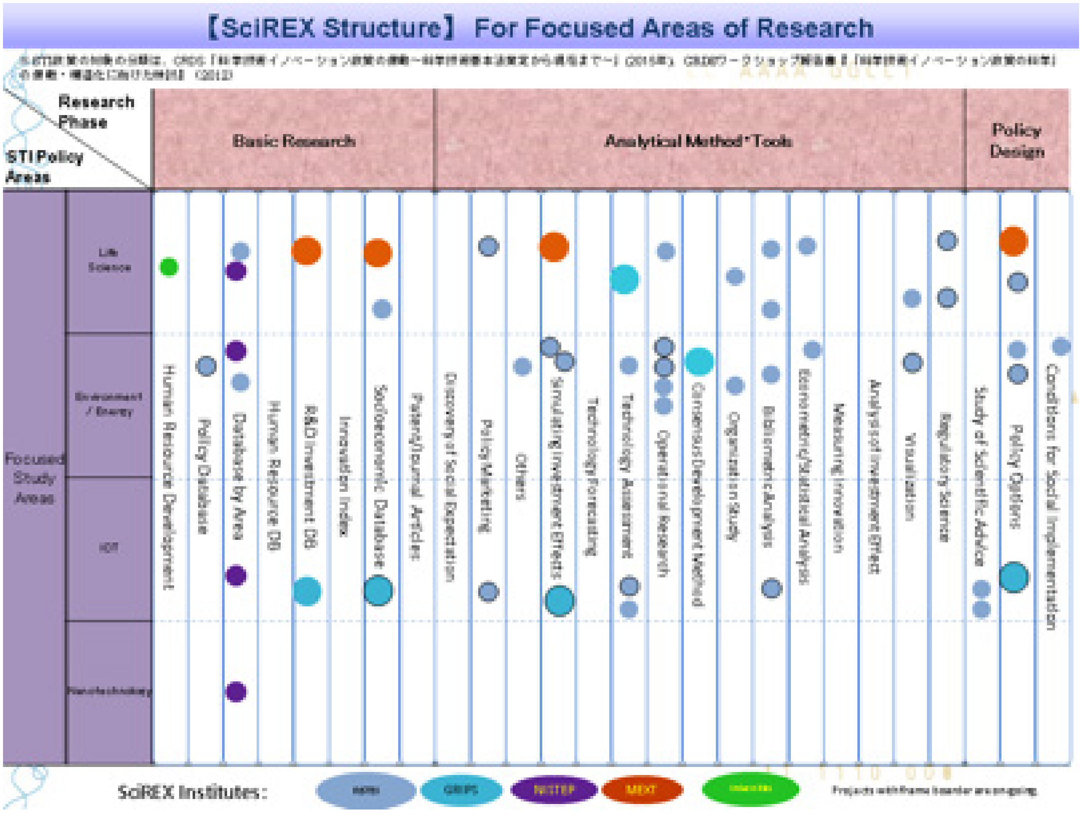
After 5 years from launching the program, it is observed some accomplishments and challenges. For defining progress of the program, we created a distribution map of SciREX activities to review management of the program as part of Structuralization. We set the activities into three categories (Resource Infrastructure - Analytical Method - Policy Design) toward policy implementation in horizontal axis. In vertical axis, we also categorize 9 domains to assort for fitting policy channels (human resources, intellectual property, research infrastructure, Evaluation Systems, Society and S&T, etc.) as well. In our analysis, by examining characteristics of the research and volume of the fund, it is determined to put a point at a certain geographical position in the two dimensional map. Having covered all activities of SciREX program, we found that the program has strong tendency of R&D investment analyses on societal and economic point of view which in a bloc of Analytical Method and Evaluation Systems. Also, the consensus development method amongst stakeholders is in area of Analytical Method and Society and S&T. Hub Universities for Fundamental Research and Human Resource Development, conducting interdisciplinary studies as well as educational courses, developed own core curriculum and provide unique activities in a bloc of Resource Infrastructure and Human Resource. With respect to research infrastructure, likewise in European Union and the United States (Guthrie et al., 2013), SciREX is currently working on harmonizing existing various research infrastructures and newly developed ones to support her studies for scientific evidence; yet, the linkage of infrastructure should be encouraged in our observation.

As mentioned, co-evolution of Science of STI Policy and STI Policy system is crucial; therefore, SciREX program is deeply committed to encourage frequent communication between science community and policymakers. For instance, last year the program held wide variety of 16 seminars which invited researchers of interdisciplinary studies as speakers. The seminar series were well functioned since stakeholders from policymakers, academia, and private sectors joined and discussed how the program can be more successful. Furthermore, so-called "Policy Liaisons", who have career background of both policymaking and scientific research, were appointed as a channel for capturing policy demand and for delivering evidence to policymaking. Nevertheless, recent survey on SciSIP community in the United States shows that research outcomes do not notably match what the policymakers expect (Sen, 2015). Although the management of SciREX program was favorably evaluated by the last year's interim appraisal by external audit, many stakeholders of SciREX share empathy with the result of the U.S. survey.

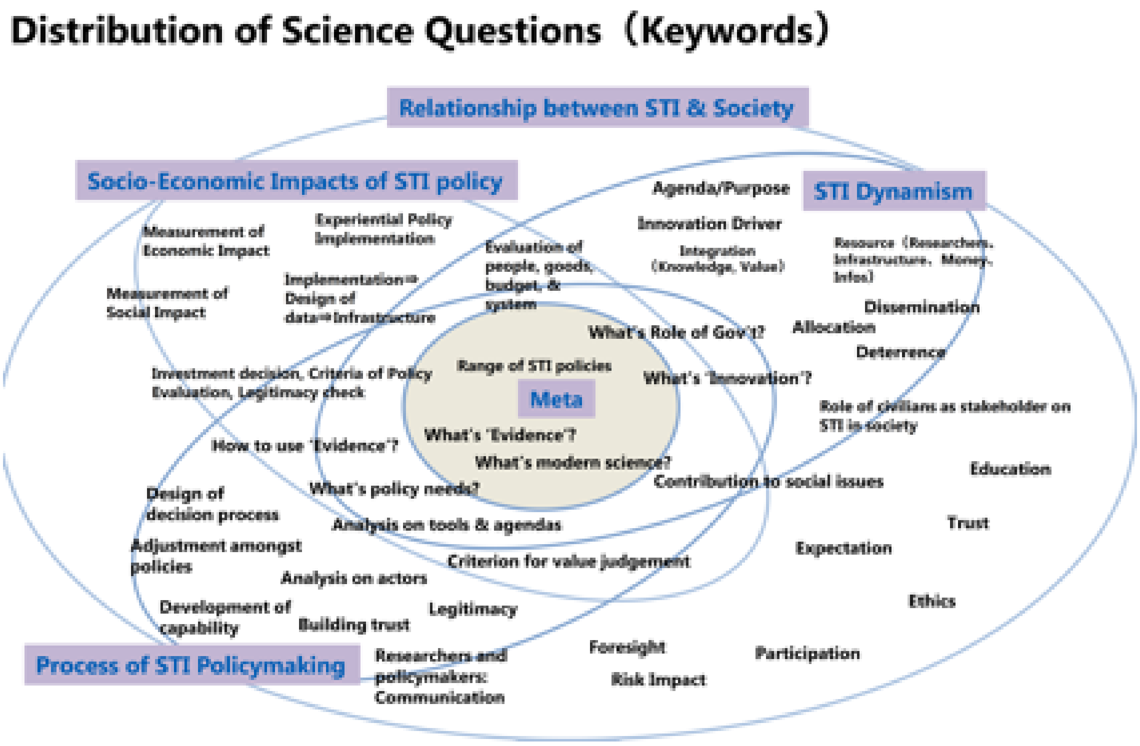
To evaluate and organize the program more efficiently, SciREX organized several meetings from beginning of 2016. The efforts, recognizing common Science Questions for improving the program, enhanced communication between policymakers and scientists. The questions become new key for evaluation of the program by linking activities and the science questions. By doing so, it turns much easier to overlook entire structure of SciREX program and to define lack of activities, which is used to be unobservable. As of November 2016, SciREX is launching to organize 'Core Curriculum' with universities in cooperation connecting to the Science Questions. Simultaneously, they plan to write a handbook for the Science of STI policy. Connecting Science Questions and distribution map of SciREX activities will present objective analysis of management of the program.

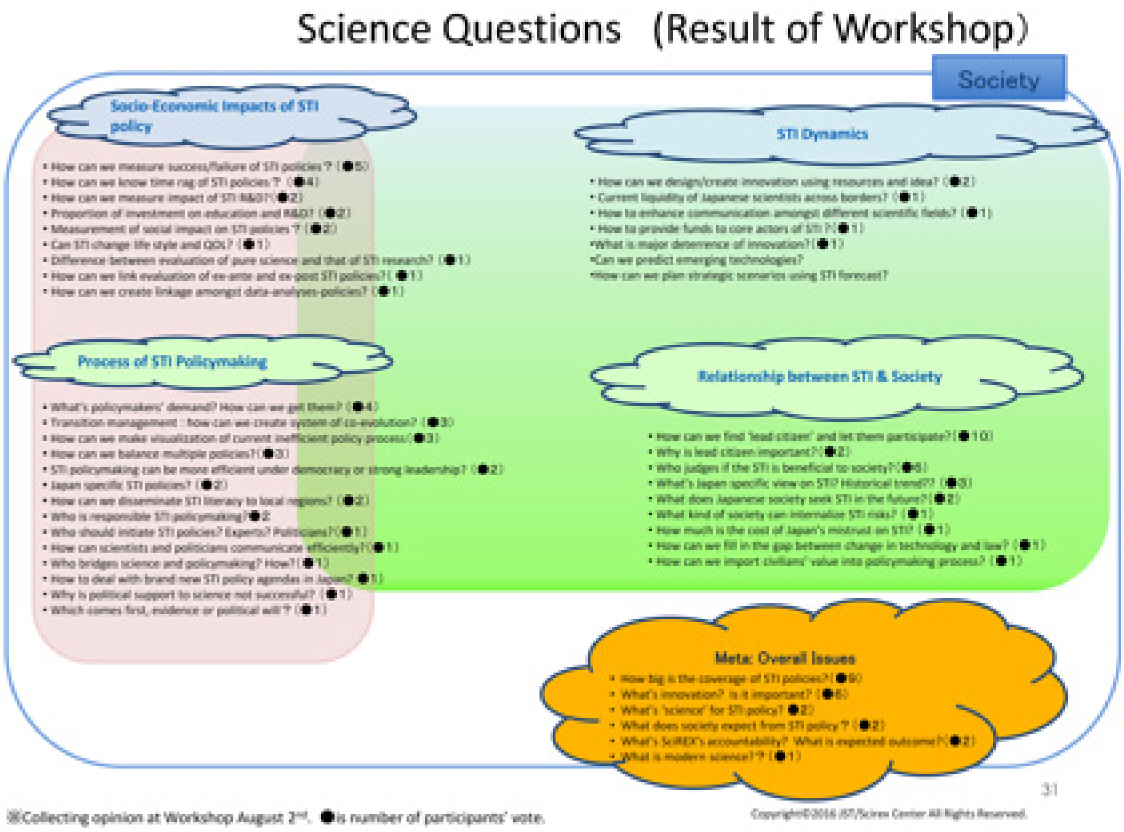
The analysis gives relevant insights into possible ways forward to harmonizing datasets, creating knowledge, and enabling innovation to achieve SciREX goals. Having analyzed development of the program, it is still necessary to consider some actions to make it sufficient system for policymaking. The sustainable incentive and strong commitment to the program is essential for participating institutes, universities, and stakeholders, to make collaborative studies and efficient networking successful. It is consistent and major challenge for SciREX program to make the initiative real 'Science' and evidence-based policymaking.





Distribution Maps of SciREX Activities as of January 2016





Results of Stracturalization Workshop as of September 2016

REFERENCE

Guthrie, Susan., Wamae, W., Diepeveen, S., Wooding, S., and J.Grant. "Measuring Research: A Guide to Research Evaluation Frameworks and Tools" Rand Europe. 2013. http://www.rand.org/pubs/monographs/MG1217.html

JST-CRDS, "Towards Realization of Evidence-based Policy Formation: Development of Science of Science, Technology and Innovation Policy", 2010. CRDS-FY2010-SP-13.
http://www.jst.go.jp/crds/pdf/2010/SP/CRDS-FY2010-SP-13.pdf (In Japanese, Executive Summary is in English).

Teich, A.H., and Feller, I. "Toward a Community of Practice: Report on the AAAS-NSF SciSIP Grantees Workshop" 2009. Washington, DC: American Association for the Advancement of Science. http://www.aaas.org/sites/default/files/scisip_report_2009.pdf

Sen, Avery., ‚SoSP: Are We There Yet? , Avery Sen. Presentation at 2015 AAAS Forum. http://prezi.com/mmvsh7kff5ao/?utm_campaign=share&utm_medium=copy

SciREX Symposium „Toward Co-evolution of Science for Policy and Policy Making Process in Science, Technology and Innovation Policy - Challenges and Future Perspectives" 2015.
http://www.scirex2015.tokyo/programEG.html

AUTHORS

SHINANO HAYASHI
Center for Research and Development Strategy,
Japan Science and Technology Agency.
K's Gobancho, 7 Gobancho, Chiyoda-ku,
Tokyo, 102-0076
Japan
shinano.hayashi@jst.go.jp
Phone: +81-3-6261-2118

TATEO ARIMOTO
Center for Research and Development Strategy,
Japan Science and Technology Agency.
K's Gobancho, 7 Gobancho, Chiyoda-ku,
Tokyo, 102-0076
Japan

National Graduate Institute for Policy Studies
7-22-1 Roppongi, Minato-ku,
Tokyo 106-8677
Japan

MASAHIRO KURODA
Center for Research and Development Strategy,
Japan Science and Technology Agency.
K's Gobancho, 7 Gobancho, Chiyoda-ku,
Tokyo, 102-0076
Japan

National Graduate Institute for Policy Studies
7-22-1 Roppongi, Minato-ku,
Tokyo 106-8677
Japan

KEIKO MATSUO
Center for Research and Development Strategy,
Japan Science and Technology Agency.
K's Gobancho, 7 Gobancho, Chiyoda-ku,
Tokyo, 102-0076
Japan

HIROAKI HARADA
Center for Research and Development Strategy,
Japan Science and Technology Agency.
K's Gobancho, 7 Gobancho, Chiyoda-ku,
Tokyo, 102-0076
Japan

RESEARCH PERFORMANCE BASED FUNDING SYSTEMS IN EUROPE

KOEN JONKERS, THOMAS ZACHAREWICZ, BENEDETTO LEPORI AND EMANUELA REALE

INTRODUCTION

Research performance based funding (RPBF) is defined as the allocation of organisational level funding to research organizations based on the (ex-post) assessment of their research performance (Hicks 2012). It is considered as one of the central tools through which many EU MS have tried to increase the effectiveness and performance of their Public Sector Research systems over the past decade.

This paper aims to analyse the extent to which RPBF allocation mechanisms are being implemented in Europe, identify strength and drawbacks of different approaches and provide an assessment of the impact RPBF systems have on research outputs of national research systems.

To do so, this study builds on a novel set of data on project and organisational level funding developed for the European Commission, which identifies funding allocation mechanisms in each of the EU-28 Member States. This approach allows to compare the scope of RPBF systems across European countries. Further, the paper build on an in-depth analysis of RPBF implementation in 28 European countries, which comes to a classification of different types of RPBF implementation. The analysis furthermore identifies a number of good practices while highlighting the potential for adverse effects of RPBF systems

DEFINING RESEARCH PERFORMANCE BASED SYSTEMS

Public research funding is generally allocated in two main ways, through project funding and through organisational or institutional level funding (Lepori et al., 2007; van Steen, 2012). This definition has been operationalized in a series of statistical projects and data are now routinely collected by EUROSTAT at national level for a number of countries. Organisational level funding for R&D can be allocated in different ways, based on historical considerations or negotiation between the State and the concerned institution or in a competitive manner (OECD, 2010). Building on Hicks (2012), research performance based funding systems are considered to be systems which base the allocation of organisational level funding for research (RPBF) on the basis of ex post assessments of research outputs.

Over the past decade many EU Member States have implemented RPBF systems, though the types of assessments and the share of resources allocated in this way differ widely. Many countries use a funding formula which is partially based on the quantitative assessment of research outputs. Another set of countries rely instead on evaluations of research output through peer review. A subset of the latter allows these peer reviews to be informed by quantitative assessments of research outputs.

The following questions will be addressed in the paper:

- How is performance assessed ex-post? The existing literature suggests focusing in this respect on the distinction between metric-based systems and systems based on peer review (possibly informed by quantitative indicators)?
- What is the nature of the link between performance assessment and allocation of resources? A major distinction, in this respect would be between an automatic relationship (through a formula) and a discretionary relationships (for example through performance contracts)?
- What is the amount of resources allocated through competitive organisational level funding of which RPBF is a major subset?

METHODOLOGY AND DATA SOURCES

The approach to analyse the nature of organisational level funding allocation systems is twofold.

First, the scope of RPBF will be delimited through quantitative data collected in the context of the EC funded PREF study. This study has engaged in the systematic collection of statistical data from national budgets, other administrative data and the accounts of research funders to assess the relative share of project funding and organisational level funding. Importantly, PREF will provide a fine-grained division of organisational level funding, which should allow distinguishing RPBF-funding streams from other organisational level funding streams.

Second, the paper will focus on a more qualitative assessment of the modalities of Performance Based Funding in the Member States. Through the information provided by a network of experts in the EU Member States, associated countries and selected third countries, 35 national research funding allocation mechanisms were examined to assess to which extent these countries implement RPBF systems.

The variables taken into consideration include: education metrics, the use of historically based funding allocation, bibliometric indicators as well as other formula elements. Other indicators frequently used refer among others to patents, external funding generated by contract research for companies or public administrations, income from Knowledge Transfer activities.

The paper then considers a number of bibliometric research output to provide some information on the level of output of national research systems which could then be compared with the characteristics of their RPBF.

RESULTS

- The scope of organisational level funding
- Data on the share of project vs organisational level funding allow distinguishing between three groups of countries:
- Countries where organisational level funding is dominant, like Italy and Spain.
- Countries where organisational level funding is more important, but project funding account for a significant share of public allocation, like Netherlands and Switzerland.
- Countries where the share of organisational level and project funding are similar, like the UK, Belgium and some Eastern European countries.

On the one hand, it might be argued that the relevance of RPBF is higher in systems dominated by organisational level funding, where the competitive component of project funding is lacking; on the other hand, it should be investigated whether the RPBF and project funding are complementary or mutually supporting, i.e. those countries having high share of project funding also introduced RPBF to a larger extent.

Performance-orientation of organisational level funding

Table 1 provides a very preliminary overview of the different groupings of countries on the basis of the nature of the RPBF system they have in place.

Table 1. Characteristics of European RPBF systems

		No RPBF											Limited RPBF			Quantitative formula with Bibliometric assessment										Peer review				
Country		BG	CY	EL	ES	HU	IE	LU	LV	MT	RO	SI	AT	GE	NL	BE (FI)	BE (WA)	CZ	DK	EE	FI	HR	PL	SE	SK	FR	IT	LT	PT	UK
Education metrics		X			X	X	X	X			X		X	X	X	X	X		X	X	X				X				X	
Historical		X	X	X	X	X			X	X	X	X						X	X									X		
Bibliometrics	Publications	X				X										X	X	X	X	X	X	X	X	X	X	X	X	X	X	
Bibliometrics	Journal Impact Based	X						X								X		X	X	X	X		X				X	X		
Bibliometrics	Citation	X				X										X	X					X		X	X	X	X			
Other formula elements	PhD graduates					X	X	X					X		X	X			X	X	X	X	X		X		X			
Other formula elements	Patents							X										X	X	X			X		X		X	X		
Other formula elements	Project funding						X							X			X	X	X	X	X	X			X	X	X			
Other formula elements	Business funding						X	X												X	X	X	X	X			X	X		
Other formula elements	Gender/diversity													X		X	X													
Other formula elements	Internationalisation													X							X	X			X	X	X	X		
Peer review																										X	X	X	X	X
Performance Contracts													X		X				X	X	X									

The table categorizes the national research funding systems into four groups according to the criteria used for their allocation. The first one is composed of countries with no Research Performance Based Funding, generally based on education metrics and historical considerations. The second group consists of countries with limited RPBF systems. The third category is composed of countries relying on quantitative formulas with bibliometric assessment to allocate research funding. Finally, the fourth group presents the countries mainly assessing research performance through peer review systems.

Effects of performance-orientation

Performance based funding, providing incentives for high impact output, is likely to have some effect on the level of excellence of the output of national science systems. The nature and the criteria of the assessment on which funding allocation are based differs across countries (see also table 1) and provide different types of incentives. Apart from the potential positive effect of these incentives on e.g. the level of excellence of the national research output, there are known to be a number of potential adverse side effects. Peer review is often associated with potential conversatism, subjectivity and can be relatively expensive (Geuna and and Piolatto, 2015). Systems that rely on bibliometrics can e.g. incentivize gaming and sub-optimal publication behaviour.

The available evidence on the effect of the different types of Performance Based Funding Systems is mixed. We find that most European systems have increased their performance on the biblometrics indicators considered. Since most public research budgets have remained relatively stable over the past decade, this is likely to be due to other factors. Potential explanations may include the growing Europeanisation or Globalisation of scientific fields (Nedeva and Wedlin, 2015). Institutional changes, including the introduction of RPBF are also likely to have played a role. There are systems without a clear RPBF system which perform very well on the bibliometric indicators considered. These systems tend to have followed institutionally rooted historical scientific development trajectories building on decades of sustained and stable funding and gradual co-development of science, higher education and industrial development. By contrast all the EU Member States which did not experience a consistent improvement in impact scores over the decade studied, did not have a RPBF system in place. These countries received recommendations by international organisations to introduce RPBF systems in recent years. There may however be other explanations for this relative under performance including chronic underfunding and the mobility of many of their best scientists to Western Europe and the US.

DISCUSSION AND CONCLUSION

This paper presents for the first time a comprehensive overview of the relative share of project and organisational level public funding for research in all EU28 Member States.

It explores the different ways in which European Member States have implemented performance based funding regimes. The European research systems can be grouped in four categories according to the type of performance based funding they have implemented: a group of countries without RPBF, a group of countries with limited RPBF, a group of countries in which the RPBF systems uses formulae based on quantitative indicators and a group of countries in which the RPBF system uses formulae based on peer review.

Factors which are likely to influence the relative effect, acceptance and success of Performance Based Funding regimes include the share of organisational level funding which is allocated through RPBF, the speed within which the system is introduced, the degree of stakeholder involvement, the impact different systems have on the autonomy of research performers, the criteria on which they evaluate and their likely impact on research excellence indicators as well as the other missions and behaviours which the government wants to promote in these organisations.

REFERENCES

Council of the European Union (2011), Council conclusion on the modernisation of higher education, retrieved 14/03/2016 from: http://www.consilium.europa.eu/uedocs/cms_data/docs/pressdata/en/educ/126375.pdf

Cruz-Castro, L., Bleda, M., Derrick, GE., Jonkers, K., Martinez, C., Sanz-Menendez, L., 2011. Issue brief.: Public sector research funding. OECD Innovation Policy Platform, retrieved 14/03/2016 from https://www.innovationpolicyplatform.org/sites/default/files/Public%20Sector%20Research%20Funding_0_0.pdf

European Commission (2011), Supporting growth and jobs an agenda for the modernisation of Europe's higher education systems from: http://ec.europa.eu/education/library/policy/modernisation_en.pdf

Geuna, A., Piolatto, M. (2016), Research assessment in the UK and Italy: Costly and difficult, but probably worth it (at least for a while), Research Policy, 45 (1), 260–271

Hicks, D., (2012). Performance-based university research funding systems, Research Policy, 41, (2), 251–261

Lepori, B., Benninghoff, M., Jongbloed, B., Salerno, C. & Slipersaeter, S., (2007). Changing models and patterns of higher education funding: Some empirical evidence. In A. Bonaccorsi & C. Daraio (Eds.) Universities and Strategic Knowledge Creation. Specialization and Performance in Europe (pp. 85-111). Bodmin, Cornwall: MPG Books Limited.

OECD, (2010). Performance-Based Funding for Public Research in Tertiary Education Institutions, Workshop Proceedings, Paris: OECD publishing

Steen, J. v., (2012). "Modes of Public Funding of Research and Development: Towards Internationally Comparable Indicators", OECD Science, Technology and Industry Working Papers, 2012/04, Paris: OECD Publishing. http://dx.doi.org/10.1787/5k98ssns1gzs-en.

AUTHORS

KOEN JONKERS
European Commission, Joint Research Centre, Institute for Prospective Technological Studies (IPTS), C/ Inca Garcilaso No. 3
E-41092 Seville
Spain
koen.jonkers@ec.europa.eu

THOMAS ZACHAREWICZ
European Commission, Joint Research Centre, Institute for Prospective Technological Studies (IPTS), C/ Inca Garcilaso No. 3
E-41092 Seville
Spain

BENEDETTO LEPORI
Faculty of Economics, Università della Svizzera italiana
via Lambertenghi 10a,
6904 Lugano
Switzerland

EMANUELA REALE
Research Institute on sustainable Economic Growth IRCRES, National Research Council CNR,
Via dei Taurini 19,
00185 Rome
Italy

HOW DO INNOVATION AGENCIES EVALUATE AND SELECT PROJECTS?

A COMPARISON OF 12 EUROPEAN AGENCIES

PETER BIEGELBAUER, THOMAS PALFINGER AND SABINE MAYER

Evaluation happens not only on the policy level, it is also an important function of innovation agencies, i.e. applied research funding organisations. Research funding agencies - regardless of focusing on applied or basic research - have to evaluate project proposals in order to select the most promising proposals for funding (Lepori et al 2007). Since the funding of societally and economically relevant research is the most important task of research funding agencies, project selection is the very core of their business.

Besides some research on peer reviewing (e.g. Lamont 2009, Mallard et al 2009, Bulathsinhala 2014, Sattler et al 2015), there is only little verified knowledge available on project evaluation and selection processes (e.g. Biegelbauer/Palfinger 2016). In a recently finished study for the Taskforce Select of the European Association of national innovation agencies, Taftie, a comparison of the respective procedures of 12 European innovation agencies has been carried out.

These are Banque publique d'investissement (BPI-France), Centre for the Development of Industrial Technology (CDTI, Spain), Enterprise Estonia (EE), The Austrian Research Promotion Agency (FFG), Croatian Agency for SMEs, Innovation and Investments (HAMAG-BICRO), Agency for Innovation by Science and Technology (IWT, Flanders), which has with 2016 been renamed into Flanders Innovation & Entrepreneurship (Vlaio), Polish Agency for Enterprise Development (PARP), Project Management Jülich (PT-Jülich, Germany), The Research Council of Norway (RCN), Netherlands Enterprise Agency (RVO), Technology Agency of the Czech Republic (TA-CR) and The Swedish Governmental Agency for Innovation Systems (VINNOVA).

The 12 innovation agencies are quite different from each other. In terms of functions the innovation agencies have to fulfil, some are very broad, such as those of BPI-France, which amongst others guarantees for bank financing and venture capital, has investments and operational cycle financing alongside banking and financial institutions, engages in equity investment directly or through partner funds and supports exports. By way of comparison e.g. the Research Council of Norway is much more directly focused towards research and technological development. Also regarding their ages, the innovation agencies vary, with e.g. the PT-Jülich having been founded in 1974 and TA-CR in 2009.

The tasks of the study were the following: provide an overview of existing selection procedures of the innovation agencies taking part in the study, analyse and compare the procedures along a variety of criteria and develop recommendations on selection procedures helpful to all Taftie member organisations.

The key points of interest were selection and role of evaluators, selection criteria, ranking procedures and general process issues. A number of critical process issues were identified and ordered after three perspectives, i.e. policy, agency and customer perspective.

The 12 innovation agencies have many different funding programmes in their portfolio. 18 programmes were chosen and the key differences between the selected programmes and their selection processes characterised. The choice of programmes / funding schemes and their selection processes was based on the following premises:

- the intervention logic of a funding scheme, i.e. the way it is to have an impact on its target clientele, influences the employed selection processes. Hence, to be able to compare and learn from comparable processes, the intervention logic of the programme or scheme for which the selection process is applied needs to be similar.
- Moreover, programmes were chosen that are widespread, so every agency interested could contribute an own programme and also other agencies shall find it easy to use the results.

Finally two programme types were chosen and their selection procedures included:

- Type 1: Grant/loan schemes for R&D with business as beneficiaries. These programmes are historically amongst the first forms of business R&D funding by the state with a high funding rate and relatively little competition.
- Type 2: Grant schemes for collaborative R&D with business and research institutions as beneficiaries. Projects / programmes can be more research driven or company driven, selection procedures may vary accordingly. These programmes historically are much younger, more competitive and normally a smaller share of proposals is funded than with type 1 programmes.

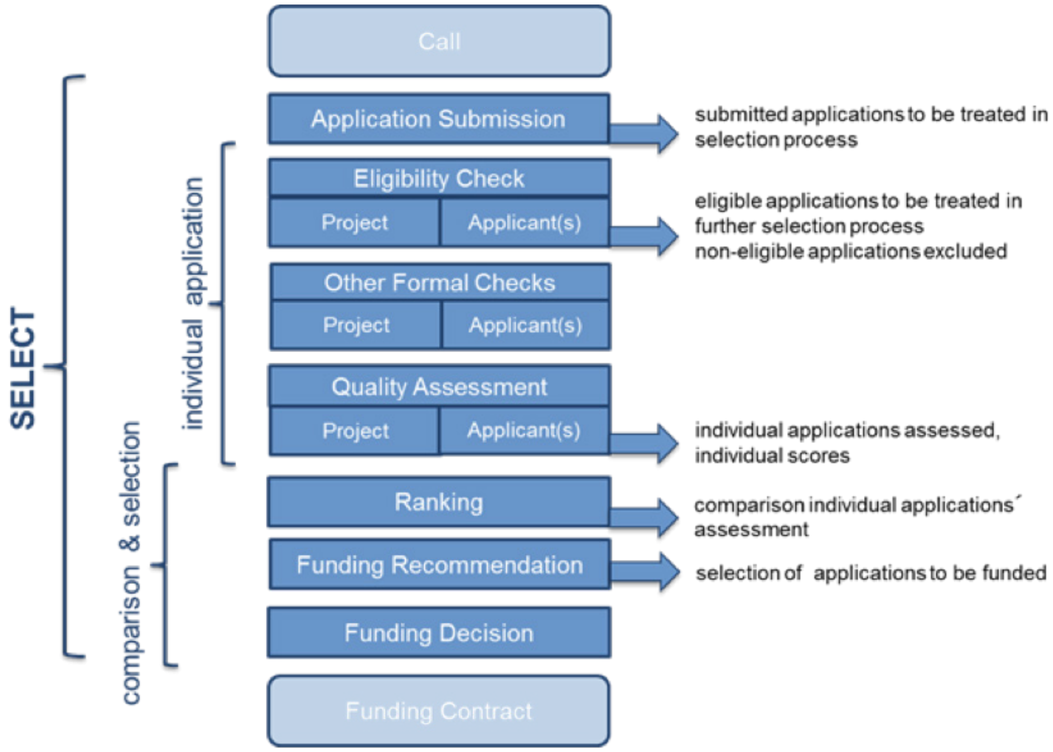
A framework was produced in order to facilitate a structured comparison against the backdrop of the challenging variety of agencies and programme types, called the backbone structure. The selection process covered here starts with the submission of the project application and ends with the funding decision. However, inputs into this process developed earlier, such as evaluation criteria, goals of the programmes, target groups for the call etc. are also covered.

Not all of the processes covered here have all the steps in place, while some will go through certain steps twice (e.g. in case of 2-step-proposals). This structure is used as a basis to describe and analyse the selected processes.

When analysing the two programme types along the backbone structure, specific characteristics become visible:

- Call (open, closed): whilst in type 1 programmes typically open calls are being used, type 2 programmes show closed calls and calls with thematic focus.
- Pre-counselling: with type 1 programmes there is typically one-to-one counselling (e.g. handling requests by firms regarding the programme), with type 2 programmes there is a concentration on information events.
- Submission: in all agencies / schemes mostly online tools are being used.

Figure 1: The backbone structure for selection processes



Source: Draft Final Report Task Force SELECT.

- Eligibility Check: both programme types use internal evaluation, in type 1 programmes sometimes applicants are directly contacted.
- Quality Assessment: with type 1 programmes more often internal evaluations (external experts mainly have tasks regarding the assessment of cutting-edge science and technology) and company visits are used. Type 2 programmes feature both internal and external evaluation but partly due to higher importance of scientific knowledge about science and technology external evaluation is more common. This circumstance leads to stronger coordination efforts within the agencies than in type 1 programmes.
- Ranking: in the selection procedures of many type 1 programmes no ranking-lists are made. In most type 2 programmes a ranking is necessary, often facilitated by a panel of experts, though there are different approaches.
- Funding Recommendation: with type 1 programmes funding recommendations more often are made by a single person (head of department, team leader etc.), whereas with type 2 programmes there mostly is a panel (selection committee, expert committee etc.), which makes the funding recommendation.
- Funding decision: There are no clear differences between type 1 and type 2 programmes regarding to the funding decision.
- Communication of funding decision: in both types of programmes applicants usually get informed by letter (or online tool). In those countries where an appeal/objection is possible applicants get more detailed information than in those countries where an appeal is either very unlikely or impossible.

Indeed, by way of comparison it becomes obvious that the differences between the practices utilised in the various agencies is sizeable, yet the differences between the procedures employed for selecting projects between the different programme types looms larger.

A major outcome of the study was the realisation that in hindsight of the differences between the agencies, their regulatory, budgetary and governance environment and the functions they have to fulfil in the respective innovation systems, it does not make sense to define a "best practice" for the selection processes (compare also Lundvall/Tomlinson 2001). Indeed, the latter have to be optimised regarding specific goals in order to be capable of speaking of "best" practices proper. They have to answer the question, "best for what?" or "best in relation to which goals?"

Rather we decided to aim for a set of "good practices" covering the project selection of innovation agencies. Accordingly, we want to define a good practice as a way of fulfilling tasks, which are understood to be effective and/or efficient in pursuing defined goals, such as producing innovations or enhancing the cooperation between firms and universities.

In fact, it soon became obvious that the innovation agencies, when selecting project proposals, have to make a number of choices. These have to be made in lieu of specific trade-offs, a few important of which shall be discussed here:

1. A decision on a very general level pertains to the form of calls to be utilised as part of the programme: should it feature closed or open calls. Accordingly, in the first case the project selection procedures will include a ranking with a competitive evaluation, whereas in the second case they will be based on single proposal evaluation on a first-come, first serve, basis. This also differentiates the two involved programme types. The distinction is caused by specific programme goals and availability of funds.

2. A further choice has to be made regarding the usage of internal and external experts in the project selection process. Both types of experts have their strengths and weaknesses (Kaufmann 2013).

   2.1 Internal expert usage may be preferred because of an expectation that they shall more strictly adhere to issues of confidentiality than external experts. The latter, however, may strengthen trust in the agency's procedures and legitimise the organisation vis-a-vis its target community.

   2.2. Confidentiality, however, usually stands in the way of transparency, therefore marking another trade-off.

   2.3. Internal experts engage more frequently into evaluation processes and therefore have often more experience, while external experts will be closer to latest developments in science and technology.

3. Organisations have to choose between efficiency and effectiveness.

   3.1. In general there is a choice between the costs of decision-making and reliability of selection procedures. The usage of several experts (e.g. four eyes principle) or invitation of highly trained experts is more expensive than less reliable practices with smaller numbers and/or less well trained experts.

   3.2. Other features of selection processes driving up its overall cost are for example efforts to standardise evaluator opinions, which may feature e.g. dominant usage of high scores or a prevalence of utilisation of low scores either due to personal idiosyncracies or cultural differences. Other evaluators might have a tendency to rate proposals higher in their own field of interest or yet others may rate those proposals lower not utilising their own preferred methodology.

4. A different form of trade-off is the tendency of many programmes to foster middle-of-the-road research using standard approaches. This may be fostered by the crowding out of evaluators, which often reason against the mainstream opinion in panel discussions, which have the task of creating a consensus between (internal or external) experts.

5. Yet another organisational choice has to be made between the evaluation of project excellence and considerations on a systemic level. There might be a trade-off between the emphasis on excellence in science and technology in a specific project proposal versus portfolio considerations aiming at the programme goal related spread of chosen projects, e.g. regarding the availability of specific technologies. Along similar lines regional aspects may be responsible for a certain project portfolio, aiming at the specific regional spread of chosen projects.

The comparison of the ways in which the 12 innovation agencies evaluate and select projects therefore shows that there is more than one solution to the challenge of financing the best research projects – "best" relating to fulfilling the programme goals. The regulatory, budgetary, socio-economic and political framework conditions the innovation agencies find themselves in form their potential options for possible and sensible solutions in the respective innovation systems. This is true for older programmes, such as type 1 schemes focusing on the competitiveness of firms, but also newer programmes, such as type 2 schemes influenced by the more societal problem oriented Grand Challenge rationales.

REFERENCES

Biegelbauer, P., S. Mayer, T. Palfinger (2016). Final Report Task Force SELECT. Vienna.

Biegelbauer, P., T. Palfinger (2016). „Auswahlverfahren von Forschungsförderungsorganisationen im internationalen Vergleich." fteval Journal 41(March) 2016, pp. 63-68.

Bulathsinhala, N. (2014). "Ex-ante evaluation of publicly funded R&D projects: searching for exploration." Science and Public Policy 42(2): 162-175.

Kaufmann, P. (2013). FTI Gutachtersysteme im Vergleich. Wien, KMU Forschung Austria.

Lamont, M. (2009). How Professors Think: Inside the Curious World of Academic Judgment. Cambridge, U.S., Harvard University Press.

Lepori, B., P. van der Besselaar, M. Dinges, B. van der Meulen, B. Potì, E. Reale, S. Slipersaeter and J. Theves (2007). "Indicators for Comparative Analysis of Public Project Funding. Concepts, implementation and evaluation." Research Evaluation 16(4): 243-255.

Lundvall, B.-Å. and M. Tomlinson (2001). Policy Learning Through Benchmarking National systems of innovation - Learning by Comparing. Report for the Advanced Benchmarking Concepts (ABC) Project. Aalborg University, Denmark.

Mallard, G., M. Lamont, J. Guetzkow (2009). "Fairness as Appropriateness: Negotiating Epistemological Differences in Peer Review." Science, Technology & Human Values 34(5): 573-606.

Sattler, D., P. McKnight, L. Naney and R. Mathis (2015). "Grant Peer Review: Improving Inter-Rater Reliability with Training." PLOS ONE 10(6).

AUTHORS

PETER BIEGELBAUER
AIT Austrian Institute of Technology, Department Innovation Systems, Vienna, Donau City Strasse 1, 1220 Vienna, Austria
tel +43-664-883-900-33

THOMAS PALFINGER
AIT Austrian Institute of Technology, Department Innovation Systems, Vienna

SABINE MAYER
Austrian Research Promotion Agency (FFG), Strategy Unit, Vienna

INSTRUMENT-SPECIFIC EVALUATION METHODS OF TEKES ACTIVITIES

JARI HYVÄRINEN

Tekes is the most important publicly funded expert organisation for financing research, development and innovation in Finland. Tekes boosts wide-ranging innovation activities in research communities, industry and service sectors. Tekes promotes a broad-based view on innovation: besides funding technological breakthroughs, Tekes emphasises the significance of service-related, design, business, and social innovations.

Tekes works with the top innovative companies and research units in Finland. Every year, Tekes finances some 1,500 business research and development projects, and almost 600 public research projects at universities, research institutes and universities of applied sciences. Research, development and innovation funding is targeted to projects that create in the long-term the greatest benefits for the economy and society. Tekes does not derive any financial profit from its activities, nor claim any intellectual proprietary rights.

The underlying foundation for evaluation is the impact model, which facilitates the identification and verification of appropriate impact mechanisms, i.e. causality between these four levels: what are the appropriate funding instruments and how should they be implemented (schemes, programmes, etc.) in order to enhance private investments in R&D (input additionality), leading to better quality R&D with higher spill-over effects (behavioural additionality), leading to more innovative products, services and businesses (output additionality) and eventually leading to better competitiveness, economic growth and innovative solutions for societal challenges (socio-economic impact).

Evaluation of R&D and innovation funding with multiple objectives and several different – often sector specific – impact mechanisms requires the use of multiple complementary methods. Some of these are qualitative and heavily based on data. However, especially in-depth learning about impact mechanisms also requires qualitative methods. Deepening this understanding also requires the continuous development of new methods, which is something Tekes has also invested resources through funding independent impact assessment research.

Evaluation has two equally important objectives: (1) to verify the real impact of R&D and innovation funding, and (2) to facilitate learning and in-depth understanding of how this impact is created and how it may be improved either by redesigning existing instruments and support measures or designing new ones. While the former has become increasingly important in order to demonstrate the value of R&D and innovation funding, Tekes places even higher importance on the latter. Learning and in-depth understanding is the key to identifying where and how R&D and innovation funding can produce the highest possible impact, thus allowing the necessary redirection or even reduction of funding necessary to reach the desired objectives.

My paper presents how the Tekes impact model will be reorganized to measure new impact goals and instrument-specific pathways. These goals take into account new insights considering the impacts of R&D and innovation funding on the whole economy and society in the Finnish innovation environment. These goals are 1) globally competitive innovative firms and economy, and 2) highly attractive innovation environment which determine the impact analysis in Tekes. The main question of the paper is how specific logic models can be formed to measure goals, instruments, beneficiary segments and industrial sectors separately?

My paper concerns two evaluation questions:

1. What are methods, results and outcomes of the Tekes activities from the perspectives of several pathways of goals, instruments, beneficiary segments and innovative sectors?
2. What are the working methods to reach new goals? What is the value added of statistical analysis, case studies, innovation-specific pathways, and meta evaluations?

The aim of public R&D investments and other actions that improve the national innovation environment is to increase wellbeing throughout society.

By evaluating the impact of innovation activities and applying metrics, we are able to establish the benefit experienced by companies receiving public RDI funding, as well as by the business sectors in question as a whole and the national economy. These kinds of evaluation are increasingly important, as the efficiency requirements set for public funding are becoming more stringent. Moreover, the findings of these evaluations, as well as predictive evaluations and projections, are used to a greater extent than before in the design and development of innovation policies.

Tekes efforts in achieving these objectives are monitored through impact analyses and reports. Tekes aims for direct well-being impacts based on innovations, productivity, structural reform and growth. The expertise and knowledge developed will extend beyond individual projects and generate the desired results over the long term.

Business renewal is particularly important from the perspective of Finland's competitiveness. The continuous development of knowledge, expertise and competencies makes an impact on the state of the national economy and the country's innovation environment. People's wellbeing have a far-reaching effect on sustainable development and new business opportunities.

When goals are so broad, it is valuable to build up more specific evaluating tools to analyze the effects of Tekes activities in parallel with Tekes impact goals. I emphasize several improvements by combining the

Tekes impact goals and the steps of impact model (inputs and resources, activities, results, and impacts on economy and society) in order to improve evaluation quality and evaluation tools focusing on impact goals, instruments, beneficiary segments and innovative sectors.

AUTHOR

JARI HYVÄRINEN, PHD
Tekes - Finnish Funding Agency for Innovation
P.O.Box 69, 00101 Helsinki
Finland
tel: +358 29 505 5803
mobile: +358 50 5577803
jari.hyvarinen@tekes.fi
Twitter: @hyvarinen_jari, www.tekes.fi

INTERNATIONAL PRACTICES OF AGRICULTURAL RESEARCH IMPACT ASSESSMENT

LAURENCE COLINET, ARIANE GAUNAND AND PIERRE-BENOÎT JOLY

CONTEXT

Agricultural research is recently facing new and broader challenges such as confronting increased competition for alternative uses of finite land and water resources, adapting to climate change, and contributing to preserving biodiversity and restoring fragile ecosystems (Interagency, 2012). But these new challenges are tougher to reach and require that Agricultural Research and Innovation System (ARIS) support great transformations. Still, recent trends are pessimistic regarding their ability to support such transformations (World Bank: Burch et al., 2007). On top of the recent concerns for ARIS, three elements further obscure the horizon. First, climate change may severely affect crop yields. Second, it is very likely that a growing part of the R&D effort will be devoted to maintenance research for maintaining high yielding production based on limited resources. Third, increasing agricultural productivity in today's context will require gains among a large number of diverse smallholders, thus reducing the economies of scale in research allowed by standardized solutions. Moreover ARIS are growingly complex and fragmented (EU SCAR, 2015). In most countries, a relative increase of private R&D is observed, that induces a strengthening of property rights, a stagnation of public R&D, and an increase in public-private partnerships. In this changing context, RIA is increasingly difficult to perform because impact cannot be attributed to a single contributing stakeholder anymore, and impact data are growingly scattered among stakeholders.

Against this background, RIA is being given a greater and renewed role. Global institutions consider that, in order to fulfill its promises, agricultural research needs a focus on challenges, an impact orientation and an improved responsiveness. Agricultural Research Impact Assessment (ARIA) has also to take into account a diversity of dimensions related to current challenges, beyond productivity gains: environmental, social impacts, impacts on food safety and occupational health.

This context is not peculiar to agriculture. In other sectors of research too, there is a strong search for methodologies of RIA that take into account broader impacts (Bozeman and Sarewitz, 2011) and improve knowledge on impact-generating mechanisms, notably through contribution analysis (Joly et al., 2015). This prompted a revival of interest in RIA methodologies, and has been the motivation for a number of projects such as: Assessments of the impacts of the Advanced Technology Program (ATP) (Ruegg and Feller, 2003), Public Value Mapping (Bozeman, 2003), the Payback Framework (Donovan and Hanney, 2011), and the Social Impact Assessment Method (SIAMPI) (Spaapen and Van Drooge, 2011). Public sector Research Organizations (PROs) are experimenting with new ways to assess the impacts of their research. PROs dedicated to agriculture are contributing to this rich field of experimentation. Practices actually implemented by agricultural PROs are poorly documented in the literature, but some authors (Khakee, 2003; Lee, 2006) claim that there is a gap between research and practice of evaluation in the field of policy evaluation.

OUR RESEARCH

This paper will give an original and updated insight on the motivations, the theoretical issues and the implementation challenges of five international agricultural public research organizations. This paper will thus enlighten the gap between the theoretical background for impact assessment that is extensively described in the literature (Bornmann, 2013) and actual practices of agricultural PROs.

A limited number of international surveys (Digital Science, 2016; Langfeldt and Scordato, 2015) analyze and compare existing practices of (agricultural) PROs in terms of their global experience in RIA. NIFU, 2015, performed such an analysis on five non-agricultural PROs. In this paper, we aim at providing original data on ARIA in practice in order to grasp the current situation and to see how the gap between academic research and practices is being dealt with. We focus on the practices of five public research organizations, selected either for their long established practices in impact assessment (Commonwealth Scientific and Industrial Research Organization -CSIRO- in Australia, and the Brazilian agricultural research organization -EMBRAPA), or because of their focus on evaluating programs (the U.S. Department of Agriculture –USDA, and the Consultative Group on International Agricultural Research –CGIAR), or for their recent intense effort devoted to research evaluation (the French National Institute for Agricultural Research –INRA). This study is of limited scope and does not pretend to be representative of what goes on in other institutions. But we suggest that, because the observed organizations are very different, our study allows catching a wide range of practices and interests for monitoring and impact evaluation.

The insight provided in this paper arises from original qualitative data that were collected through interviews conducted with senior managers of each PRO (1 to 2 informants by PRO), and desk research based on resources from the organizations' web pages and internal management documents passed on by our informants.

After a brief linear description of each PRO and its history of RIA, a cross-cutting analysis is performed in order to compare their purposes while performing RIA, the way RIA is designed, implemented, and the way its results are actually used.

RESULTS

PURPOSES, PERIMETERS AND CALENDAR FOR RIA

It appears that all five public research organizations (PROs) assign multiple purposes to their impact assessment approach, usually accountability to funders, organizational learning, and internal capacity building. Accountability objectives usually requires an external evaluation or validation, which can go against the objectives of involving internal staff and develop internal evaluative capacities, and can even prompt researchers to report overestimated data about their impact thereby limiting internal organizational learning. The value of the evaluation system can therefore be characterized according to the balance between these three — sometimes-divergent — objectives.

Impact assessment may concern multiple levels of the organization: national or international research programs (CGIAR, CSIRO, USDA), scientific divisions, centers, or business units (CSIRO, CGIAR, INRA), the organization as a whole (CSIRO, INRA, EMBRAPA, USDA).

In terms of rhythm, the approaches are designed to be inserted in external assessment schedules. RIA systems build organically on the PRO institutional structure and operate in rhythm with existing processes; they differ if funding comes through the institution or through programs.

EVALUATION DESIGN

The PROs we studied have explored in recent years new ways to evaluate their impact, and have set guidelines to standardize the way societal impact should be assessed (CSIRO being the most recent). It happens that the more recent developments are sometimes not yet implemented on a routine basis (e.g. CSIRO). Some evaluation designs are based on monitoring systems (CGIAR, CSIRO, USDA) which are integrated into a whole theory of change, creating a thread between activities, expected outputs and outcomes, and aiming at tracking progress towards expected societal impact at the end of each funding phase. This "targeted" impact assessment approach faces a challenge related to time lag since the timespan imposed for the evaluation is often too short to observe time-distant societal impact. The two Ex-post assessment approaches of our sample (EMBRAPA, INRA) do not depend on the research initial objectives, and in this sense they are goal-free.

All five organizations assess their economic impact, and most of them account for environmental and other broader social impact. In terms of methods, some PROs (INRA, EMBRAPA, and CGIAR) combine aggregated econometric approaches and case-study-based approaches, with some attempts to complement each approach by the other. All organizations (with the exception of USDA) perform case-studies although in different ways: narratives illustrating the quantification of impacts (EMBRAPA) or cost-benefit analysis (CSIRO), or case-studies encompass deep qualitative analysis of processes and networks (INRA). In any case, PROs attempt to quantify impacts, often through monetization (CSIRO, CGIAR). There are some attempts to quantify the impact with physical indicators specific to each impact dimension relying on home-made metrics (Joly et al., 2015; Rodrigues et al., 2010). The quantification of impact is often assorted with some incentives to attributing a share to the PROs research effort (CSIRO, EMBRAPA, CGIAR), which may lead to bias in case selection, in favor of more recent or less collaborative cases.

The tendency towards centralized monitoring and evaluation systems (CGIAR) or standardized guidelines for evaluation (INRA, CSIRO, EMBRAPA) answers the managers' and funders' desire to bring value of local evaluations at higher level and produce information on global impact. Aggregation is often performed at the expense of detailed meaningful information but it entails a standardized methodology to study comparable cases, which is lacking in highly decentralized implementers such as CGIAR Research Programs.

IMPLEMENTATION

All PROs are very concerned with the external credibility of their monitoring and or evaluation system, and the impact they report. Measures to promote external credibility include open calls for proposals to develop methodological supports (CGIAR), call for consultant tenders to apply standard guidelines (CGIAR, CSIRO), external validation of the case reports (CSIRO, INRA).

This survey reports that in some organizations (INRA, EMBRAPA, CSIRO), interactions between research and practice is organized in a systematic way: in these cases social scientists are involved in the design of approaches. Examples gathered from this study show however a gap between the theoretical method (Bornmann, 2013) designed by academic research, the design that PROs themselves wish to institutionalize (Walker et al., 2008 at CGIAR; EMBRAPA, 2015 ; Heisey et al. 2010 for USDA/ERS; CSIRO, 2015; Joly et al 2015 for INRA) to assess their impact, and the approach PROs actually implement. This theory-practice gap may be explained by budget constraints (CSIRO and CGIAR), calendars misalignment (ex-post assessment becoming relevant after project budget has been discontinued: CGIAR and CSIRO), management issues (staff willingness, skills and coordination: CGIAR), data availability (INRA, CSIRO, CGIAR), or time delays to implement recent management changes (CSIRO, CGIAR, INRA).

USE OF THE RESULTS OF IMPACT EVALUATION

RIA is reported to affect PROs management practices in a variety of ways, even if paradoxically, they make only limited uses of the results yielded so far. The main use that all the PROs effectively make is demonstration to stakeholders. While INRA, USDA and EMBRAPA seem to build accountability on the motivation and interest of a diversity of stakeholders, CSIRO and CGIAR prioritize reporting on the good use they make of public funds. Information is lacking on how CSIRO, CGIAR, or USDA ground their funding allocation decision on the basis of societal impact assessment. In some instances, experts reviewing research programs suggest that the funding allocation decision may be poorly served by present monitoring and evaluation systems. Similarly, institutional learning based on ex-post assessment seems limited, which is not surprising since the case-study methods used in the PROs considered do not go very far in terms of understanding of the impact generation mechanisms. Learning objectives seem to be chiefly achieved through monitoring approaches, and may concern low-level tactical topics, rather than strategic higher level issues.

ANALYSIS AND CONCLUSION

This study confirms the importance of RIA for PROs nowadays and reveals some systematic trends in the approach implemented.

All five organizations have recently made serious attempts to improve the way they evaluate their societal impact. This confirms the central importance of this development, in relation to the challenges of agricultural R&D. Credibility is important to PROs and the five organisations of our study have looked for methods that would combine excellence validated by their scientific peers and effectiveness in expressing outcomes or impacts for a specific audience of stakeholders or funders.

PROs have multiple ambitions for impact assessment. Accountability to funders (Treasury or Donors) is clearly an important driver for institutionalizing impact evaluation and monitoring systems. One area where funders exert influence is their demand for quantitative targets and metrics for outputs, outcomes and impacts (USDA, CGIAR, CSIRO, EMBRAPA). The study suggests that in such instances (CGIAR, EMBRAPA) this objective may be served better than institutional learning on how to produce impact or capacity-building benefits. However there is a gap between PROs ambitions for their RIA systems, and what they actually implement. The belief that learning and accountability goals can be mutually accommodated is widespread. In the context of niche experiments, Regeer et al., (2016) for example emphasize that "although evaluation for accountability and evaluation for learning in practice are often experienced as oppositional", different types of accountability exist (upwards for funders, downwards for beneficiaries and internal for managers) and different forms of evaluation may fulfill the variety of needs claimed. Our study suggests that, in practice, there is a tension between these objectives. From our interviews it appears that learning, and capacity-building, are two important motivations for PROs' staff (CGIAR, CSIRO, INRA) to take part in evaluation, however it is also evident that the accountability objective is of more universal interest for funders. Rhetoric is also widespread about the importance for an organization of discerning what is not working. But the question arises that, for an organization, exposing strategic flaws or weaknesses could come at a cost in terms of reputation or future funding (CGIAR, USDA). This is especially the case when there is a sense that decisions in terms of funding flows would be immediately associated with such information, which can be a common case when funds come through programs. In this situation immediate operational and tactical learning of lower importance will be more readily addressed. Systems that introduce more distance between evaluation and funding decisions (EMBRAPA, INRA) leave more opportunities for strategic learning. Considering the tensions between the multiple objectives of RIA, it appears that making strategic choices is necessary. This leads to point to differences between styles of evaluation. Power (1994) identifies two contrasted ideal-types of evaluation (type 1 or type 2), characterized by a set of dichotomies. According to Power, evaluation may be oriented toward external control (type 1) or internal learning (type 2); it may be one-dimensional or multidimensional; evaluation process may assume low trust or high trust between evaluators and evaluated; evaluation may be performed by external controllers or distributed insiders; etc. To take into consideration the characteristics of research and the importance of learning and capacity building research organizations, should foster type 2 evaluations.

REFERENCES

Bornmann, L., 2013. What is societal impact of research and how can it be assessed? a literature survey. Journal of the American Society for Information Science and Technology 64, 217–233.

Bozeman, B., 2003. Public Value Mapping of Science Outcomes: Theory and Method, in: Knowledge Flows & Knowledge Collectives: Understanding the Role of Science & Technology Policies in Development. D. Sarewitz, Et. Al., 2.

Bozeman, B., Sarewitz, D., 2011. Public value mapping and science policy evaluation. Minerva 49, 1–23.

Burch, D., Lawrence, G., Green, G.P., Ichijo, K., Nonaka, I., Pimentel, M., Bower, J., Gilbert, C., Couto Filho, V., Flavio, L., 2007. World Development Report 2008: agriculture for development. The World Bank, Washington D.C.

Digital Science, 2016. The Societal and Economic Impacts of Academic Research. international Perspectives on good practice and managing evidence (Digital Research Reports No. 978-0-9929477-5–0), Holtzbrinck Publishing Group. Digital Science, London.

Donovan, C., Hanney, S., 2011. The "Payback Framework" explained. Research Evaluation 20, 181–183.

EMBRAPA, 2013. EMBRAPA's Social Report 2012. Distrito Federal (DF): Embrapa, Communication Secretariat, Strategic Management Secretariat, Brasilia.

EU SCAR, 2015. Agricultural Knowledge and Innovation Systems Towards the Future – a Foresight Paper. EU SCAR, Brussels, Belgium.

Interagency, 2012. Sustainable Agricultural Productivity Growth and Bridging the Gap for Small-Family Farms: Interagency Report to the Mexican G20 Presidency. Interagency: Bioversity, CGIAR Consortium, FAO, IFAD, IFPRI, IICA, OECD, UNCTAD, Coordination team of UN High Level Task Force on the Food Security Crisis, WFP, World Bank, and WTO.

Joly, P.-B., Gaunand, A., Colinet, L., Larédo, P., Lemarié, S., Matt, M., 2015. ASIRPA: A comprehensive theory-based approach to assessing the societal impacts of a research organization. Research Evaluation 24, 440–453.

Khakee, A., 2003. The emerging gap between evaluation research and practice. Evaluation 9, 340–352.

Langfeldt, L., Scordato, L., 2015. Assessing the broader impacts of research. A review of methods and practices (Working Paper 8/2015). NIFU, Oslo.

Lee, N., 2006. Bridging the gap between theory and practice in integrated assessment. Environmental impact assessment review 26, 57–78.

Power, M., 1994. The audit explosion. Demos.

Regeer, B.J., de Wildt-Liesveld, R., van Mierlo, B., Bunders, J.F., 2016. Exploring ways to reconcile accountability and learning in the evaluation of niche experiments. Evaluation 22, 6–28.

Rodrigues, G.S., de Almeida Buschinelli, C.C., Dias Avila, A.F., 2010. An environmental impact assessment system for agricultural research and development ii: institutional learning experience at Embrapa. Journal of technology management & innovation 5, 38–56.

Ruegg, R., Feller, I., 2003. A Toolkit for Evaluating Public R&D Investment: Models, Methods, and Findings from ATP's First Decade. (Grant/Contract Report). National Institute of Standards and Technology, Gaithersburg.

Spaapen, J.M., Van Drooge, L., 2011. Introducing "productive interactions" in social assessment. Research Evaluation 20, 211–218.

Walker, T., Maredia, M., Kelley, T., Rovere, R.L., Templeton, D., Thiele, G., Douthwaite, B., 2008. Strategic Guidance for Ex Post Impact Assessment of Agricultural Research. Report prepared for the Standing Panel on Impact Assessment, CGIAR Science Council, Science Council Secretariat: Rome, Italy.

AUTHORS

LAURENCE COLINET
INRA
147 rue de l'Université
75338 Paris Cedex 07
France

ARIANE GAUNAND
INRA, Délégation à l'Evaluation
147 rue de l'Université, 75338 Paris Cedex 07,
France
and
Université Paris-Est
LISIS 77454 Marne-la-Vallée,
France

tel: 0033 (0)1 42 75 91 06
ariane.gaunand@paris.inra.fr

PIERRE-BENOÎT JOLY
Université Paris-Est
LISIS 77454 Marne-la-Vallée,
France

CONTRIBUTION OF SOCIAL NETWORK ANALYSIS FOR EVALUATING IMPACTS OF SCIENCE-BASED RESEARCH AND INNOVATION PROGRAM: THE EXAMPLE OF THE FARMERS' CONVERSION TO ORGANIC CROP PRODUCTION IN CAMARGUE

QUIÉDEVILLE SYLVAIN

INTRODUCTION

This paper aims to demonstrate the interest of performing a Social Network Analysis (SNA) for ex-post evaluating Impacts of Science-Based Research and Innovation program (ISRIP) in the agricultural sector. In the EU funded IMPRESA project[1] (Impacts of Research on EU Agriculture), the approach of "ISRIP Pathway Analysis" was developed to assess the role of agricultural research based innovation (Quiedeville et al., n.d.). The "ISRIP Pathway Analysis" approach is based on the Participatory Impact Pathway Analysis (PIPA); and the conduct of stakeholders' workshops (with researchers, funders, institutions, extension services, and farmers) is the guiding thread of it. The approach comprises a central workshop dedicated to the evaluation of the research program (set of projects) under review. In this workshop, stakeholders are asked to reconstruct the theory of change of the research program by identifying changes (outcomes) and defining the way they happened (via research activities, outputs, etc). The "ISRIP Pathway Analysis" approach includes a Social Network Analysis (SNA), among other complementary methods to PIPA, but the rationale of conducting a SNA needs to be further reflected. We explore this through the case of the transition to organic farming in Camargue, which was performed in the IMPRESA project as one of the six case studies conducted.

In the context of ex-post evaluating ISRIP, we made the hypothesis that SNA aids (as part of the "ISRIP Pathway Analysis" approach) to understand how new techniques or products are spreaded and thus to help draw conclusions on the impacts and role the research in the whole process. We concentrated ourselves on SNA, which is in line with the concept of innovation system (Lundvall 1992; Touzard et al. 2014) that challenged since many years the classical mode 1 of linear knowledge transfer by Gibbons et al (1994).

THE CAMARGUE CASE

The Camargue territory is situated in the south east of France over an area of some 145,000 ha. Rice is the main production cultivated and organic agriculture started in the eighteens. At that time no specific value chain was dedicated to organic products, but the trader SARL Thomas has given up to conventional farming in 1990 (it handled around 6,500 tons of rice) to actually concentrate on organic crop production. The cooperative SudCéréales also positioned itself on the organic market but marginally, and the firm BIOSUD has been founded in 2003 with the goal of organizing the organic value chain in a single common objective of negotiating and selling products through a specialized company. In 2000, the National Institute of Agricultural Research (INRA), the International Centre of Agricultural Research (CIRAD) and the French Centre of Rice (CFR) have launched a research program in order to develop organic crop production systems in Camargue. This research program was evaluated in the IMPRESA project.

METHODOLOGY

SNA INDICATORS

The SNA indicators of betweenness, clustering coefficient, density, and "degrees" were chosen to help analyze the impacts and role of the research in the Camargue case. We hypothesized that the identi-

1 For more details, refer to: http://www.impresa-project.eu/home.html.

fication of actors with a high betweenness is of particular interest, as those actors are likely to be knowledge brokers (Haythornthwaite 1996). Indeed, "the betweenness of a point measures the extent to which an agent can play the part of a "broker" or "gatekeeper" with a potential for control over others" (Scott 2000). The clustering coefficient of an actor is the quotient of its level of connectivity among its neighbors on the total possible number of connections that may occur between those neighbors. Its calculation intended to define whether the different actors are connected to a structured organization; thereby to help understand the evolution of actors' position in the whole network and whether the research has played a role in it. More generally, we assumed that the clustering coefficient can aid to estimate how resilient and robust the actor network is as well as its capacity to support innovations. The density (average number of relationships among actors) could be seen as an economic performance indicator through enhancing information flow (Vurro, Russo, and Perrini 2010). Finally, the "degrees" allow examining the evolving strength of connectivity from one actor to another; and could help to understand how the research system has contributed to the change. The table 1 summarizes how SNA data were collected and analyzed. Three steps have been followed: (1) Face-to-face interviews; (2) generalization of the sample; and (3) calculation of SNA indicators.

order to ameliorate future research programs. However this approach may lack scientific rigor if the different information gathered would not be further explored and validated by identifying clear evidences (e.g. from reliable available documents, official statistics, etc). This is why the process tracing method was applied as part of the "ISRIP Pathway Analysis" approach in the Camargue case. In a nutshell, it intends to evaluate whether the first and second event of each pathway link actually occurred; if the link can be explained by an underlying mechanism; and if the second occurrence of the link was due to other factors. This procedure also applies to pathway links specifically related to relationships issues. Given the complexity of the procedure, the "ISRIP Pathway Analysis" approach provides the opportunity of organizing all the information in a so-called "table of links" (see table 2). The origin (first event) and destination (second event) of the pathway links are specified in the first two columns, whilst the other columns relate to underlying mechanisms and alternative explanations.

Table 1: Collecting and analyzing of SNA data

SNA steps	Target(s)	Explanations
Face-to-face interviews	Researchers from INRA and CFR Respondents from private traders 11 farmers (7 partial-organic and 4 organic)	We asked for useful relationships (information flow, financial exchanges, and collaborative ties) around organic agriculture An intensity score from 0 to 3 was set Six times periods were considered over the years 2000-2014
Generalization of the sample	The population (all organic farmers, researchers, extension services, and rice traders)	We did a simple transposition of the sample of 11 farmers, which was representative, to the population (35 farmers). The interest of generalizing the sample was to ensure that stakeholders are not under or over represented in the network.
Calculation of SNA indicators	Betweenness Clustering coefficient Degrees	Calculation of the indicators by the UCINET software (Borgatti, Everett, and Freeman 2013).

TABLE OF LINKS

The "ISRIP Pathway Analysis" approach is of participatory nature and actively involves stakeholders in the evaluation process of the research program and innovation under review. The rationale of this is mainly to increase the plausibility that stakeholders will use evaluation results in

Table 2: Blank table of links

Pathway links		Description of the underlying mechanism(s)	Alternative explanations of the mechanism(s)	Validity of the alternative explanations
Origin of the link	Destination of the link			
Example: Activity 1 (name to be specified)	Example: Output 1 (name to be specified)	*Specify the most relevant evidences as to how the first event of the link has led to the second occurrence*	*Specify the plausible alternative explanations to the link*	*Yes or no* *If yes, specify its importance*

Source: (Quiedeville et al., n.d.).

RESULTS

We have done five SNA tests on the basis of stakeholder's statements. Below we summarize the four more important ones (out of five).

TEST 1:
GROWING INFLUENCE OF INRA IN THE NETWORK.

The first suggestion made was that some research activities (mainly undertaken by INRA) have led to an increase of the influence of INRA in the network. The SNA has allowed to confirm this. The betweenness score of INRA has evolved from 370 in 1999 and 415 in 2006 to 542 in 2014. Furthermore, we can confirm the hypothesized underlying mechanism. In effect, we observed growing relationships between INRA and farmers, with an increase of around 80% in their bilateral "degrees" over the years 1999-2014. That said, these bilateral relationships started to increase in the year 2005, which means that the CEBIOCA project (the first research activity done about an agronomic diagnosis) did not play a significant role. First experimentations in farming plots and the participatory training sessions have boosted the interactions between INRA and farmers. One of the alternative explanations hypothesized was the increase in relationships between the neighbors of INRA, as it could also explain the growing centrality of INRA in the network. This hypothesis was validated, as we observed a growth of 60% of the clustering coefficient[2] of INRA (from 0.1 in 1999 to 0.16 in 2014). As a result, the SNA does not fully corroborate what the stakeholders claimed in workshops. It appears that the research and disseminations activities done by INRA were not the only factors explaining its growing influence in the network around organic farming in the Camargue.

TEST 2:
INFLUENCE OF CIRAD IN THE ACTOR NETWORK.

The second suggestion made was that some research activities have led to a growing influence of CIRAD in the network. The SNA has allowed to confirm the growing influence of CIRAD within the network. During the time span of the program, the betweenness of CIRAD has increased about 34% from 1999 to 2014 and the average "degrees" around 61%, whereas the average "degrees" only increased about 29% in the entire network. Furthermore, the hypothesized underlying mechanism was also confirmed. In effect, relationships between CIRAD and farmers were growing, which is revealed by an increase in the bilateral "degrees" about 45% (from 11 over the years 1999-2010 to 16 in 2014). However two alternative explanations were confirmed. The first is the increase in relationships between CIRAD and SudCéréales as well as between CIRAD and INRA. The second is the growing interactions between the neighbors of CIRAD. This is illustrated by a growth of 60% of the CIRAD's clustering score (from 0.2 in 1999 to 0.32 in 2014). This situation raises the complexity of the innovation network and the importance of the role played by complex interrelationships among various actors.

TEST 3:
STRUCTURING OF THE ACTOR NETWORK.

The third suggestion made was that the increasing influence of both INRA and CIRAD have developed the exchanges and links in the network about transition to organic farming. The hypothesized underlying mechanism i.e. INRA and CIRAD have become knowledge brokers for the transition to organic farming, was corroborated by their higher betweenness.

TEST 4:
ADOPTION PROCESS.

The fourth suggestion made was that the structuring of the network has contributed to the adoption of organic farming and to crop rotation development (useful incremental innovation to switch to organic farming). The main hypothesized underlying mechanism was the development of information sharing between INRA and farmers, which was confirmed by the previous tests done. A main alternative explanation was the possible presence of peer-to-peer exchanges between farmers. In fact, the vast majority of the farmers could not find any relevant relationships with their colleagues as concerns organic crop production.

With respect to impacts on the organic actor network, note that we observed an increase of 44% and 50% (since 2000) of the clustering coefficient and the density, respectively.

DISCUSSION AND CONCLUSION

The SNA approach contributed successfully in the evaluation of IS-RIP. Particularly, it has allowed the different hypothesized pathway links on relationship issues to be deeply examined.

The SNA could not tell by itself what the effects of receiving information on the actors are and if their behaviors have changed and through which mechanisms. We had to make the assumption that changes in actors relationships were correlated to the evolution of the innovation. We could set this assumption since we only considered relevant relationships for organic farming.

However, SNA was very interesting for confirming or contradicting stakeholders' statements on relationships issues. Therefore we see SNA as a good way to triangulate the different information collected and increase the plausibility that we draw accurate conclusions regarding the impacts and role of the research as well as on the way the innovation pathway occurs. Finally, the SNA suggests that research on Camargue organic crop production has implied the actor network to be both more resilient and likely to support development of further innovations towards sustainable food systems.

2 Individuals with high clustering coefficients (central actors) are linked to actors who are well connected together.

REFERENCES

Borgatti, SP, MG Everett, and LC Freeman. 2013. Analysing Social Networks. SAGE Publications Ltd; 1 edition.

Gibbons, Michael, Camille Limoges, Helga Nowotny, Simon Schwartzman, Peter Scott, and Martin Trow. 1994. The New Production of Knowledge: The Dynamics of Science and Research in Contemporary Societies.

Haythornthwaite, Caroline. 1996. "Social Network Analysis: An Approach and Technique for the Study of Information Exchange." Library & Information Science Research 18 (4): 323–42. doi:10.1016/S0740-8188(96)90003-1.

Lundvall, Bengt-Ake. 1992. National Systems of Innovation: Towards a Theory of Innovation and Interactive Learning. Pinter Pub Ltd .

Quiedeville, Sylvain, Dominique Barjolle, Matthias Stolze, and Jean-Claude Mouret. n.d. "Unpublished. Ex-Post Evaluation of Impacts of Science-Based Research and Innovation Program: A New Method Applied in the Case of Farmers' Transition to Organic Production in Camargue." Food Policy.

Scott, John. 2000. Social Network Analysis: A Handbook. London: SAGE Publications.

Touzard, Jean-Marc, Ludovic Temple, Guy Faure, and Bernard Triomphe. 2014. "Systèmes D'innovation et Communautés de Connaissances Dans Le Secteur Agricole et Agroalimentaire." Innovations 43 (1): 13. doi:10.3917/inno.043.0013.

Vurro, Clodia, Angeloantonio Russo, and Francesco Perrini. 2010. "Shaping Sustainable Value Chains: Network Determinants of Supply Chain Governance Models." Journal of Business Ethics 90 (S4): 607–21. doi:10.1007/s10551-010-0595-x.

AUTHOR

QUIÉDEVILLE SYLVAIN (CORRESPONDING AUTHOR)
Research Institute of Organic Agriculture (FiBL)
Department of Economics and Social Sciences
Ackerstrasse 113
Postfach 219
5070 Frick
Switzerland

RTI EVALUATION AS GOVERNANCE AND EFFECTIVENESS TOOL: THE CASE OF EMBRAPII IN BRAZIL

SERGIO SALLES-FILHO, ADRIANA BIN, NICHOLAS VONORTAS, RAFAELA M. ANDRADE, PAULA F. D. CASTRO AND FERNANDO A. B. COLUGNATI

Research, technology and innovation (RTI) policies have become enormously diversified in the last decades around the world, as well as increasingly recognized by their contributions to economic and social development. In this context, evaluation of such policies has gained importance, *"(…) both as a policy and management supporting tool and as a tool to assess policies in order to justify or redirect funding"* (Edler et al., 2012, p. 167).

Multiple approaches have been employed to perform RTI policy evaluation in order to answer demands for transparency, accountability and performance regarding societal investments (Link and Vonortas, 2013) and thus to contribute to policy-making process. Among the several challenges regarding the choice of the best design and methodologies to perform a particular RTI evaluation, one can distinguish the need for taking into account that not only policies' results and impacts are multiple, but they are also perceived differently by multiple policies' beneficiaries.

Under this perspective evaluation is crucial to understand social dynamics (Patton, 2012) and to create governance among different *stakeholders* involved in the policy implementation (Whitley and Glaser, 2007).

This particular challenge is notably important within the evaluation of programs and instruments oriented to support collaboration for R&D and innovation and enhance the connection among different actors, such as science-industry research centers, collaborative research and programs, and collaborative knowledge exchange projects (Cunningham and Gök, 2012).

The purpose of this manuscript is twofold: to contribute to the debate on how evaluation can make STI policy design more effective, and to evaluate a specific policy instrument headed to promote innovation and R&D development highlighting governance and managerial capabilities.

For this, the manuscript presents the results of an in-deep evaluation carried out by the authors in 2015 and 2016 focusing on a new policy instrument called Brazilian Company for Industrial Research and Innovation (EMBRAPII).

Brazil, as well as other developing countries, has historically presented an enduring mismatch between scientific production and the introduction of new products, services and processes, which means that the country has been more successful in their research policies – increasing, for instance, its position in terms of scientific production in indexed databases such as Scopus and Web of Science – than in their innovation policies. Aggregate data from Brazil shows that the efforts and results from innovation processes are still far below the OECD medians.

In the past 15 years Brazil has implemented some important initiatives in RTI policies, which include (but are not restricted to) those with focus on collaboration between research organizations and universities and firms (Salles-Filho et al., 2012; Kannembley Jr. and Porto, 2012; Suzigan et al. 2009; Hochstetler and Montero, 2013). However, as pointed out by Pacheco and Corder (2010) such initiatives are characterized by lack of prioritization, insufficient resources, discontinuity and the already mentioned mismatch between research and innovation.

In an attempt to develop and implement an approach of innovation policy, a new instrument was launched in 2013. This initiative, inspired by other national models (for instance, the German Fraunhofer, the French Carnot Institute, and the Korean Kaist, among others), was implemented through the creation of EMBRAPII, a private-not-for-profit funding and governance agency working under management contract with the Ministries of Science, Technology and Innovation and of Education.

EMBRAPII's mission is to foster innovation in Brazilian industry through pre-competitive R&D projects in collaboration between companies and industrial research organizations, lowering innovation risks, following the common rationale for intervention in this kind of policy (as discussed by Cunningham and Gök, 2012).

From 2013 to 2015 EMBRAPII implemented a pilot phase that resulted in 63 R&D and innovation projects executed by three Brazilian research organizations (ROs) - Institute for Technological Research (IPT), National Institute of Technology (INT) and National Service of Industry's Integrated Campus for Manufacturing and Technology (SENAI-CIMATEC), in collaboration with 44 firms. Some of these projects are still running, with deadlines foreseen to mid-2016.

The pilot phase represented a total investment of circa US$ 50 million, being one third supported by EMBRAPII, one third by the ROs and one third by the companies. After the pilot phase EMBRAPII initiated its steady-state phase, with 13 ROs and a budget of circa US$ 350 millions to be invested until 2018.

EMBRAPII can be considered a new policy instrument in Brazil in at least three main characteristics: a) once a RO is accredited as an EMBRAPII Unit it has immediate access to funds in order to contract R&D projects directly with companies; b) companies are involved since the initial phases, presenting their demands, specifying their focus and approaches and negotiating contractual conditions with ROs; c) contracted projects are executed by ROs and monitored by companies, and a project only concludes when it receives a "letter of acceptance" from the company directly involved.

The evaluation of the EMBRAPII's Pilot Phase was oriented to identify and measure two main themes: the outputs and outcomes of the R&D

and innovation projects (technological results and its appropriation) and the behavioral changes of involved actors (following good practices of R&D and innovation planning and management). In addition, the evaluation was also concerned with the creation of governance among the beneficiaries of the policy.

To fulfill these purposes, data collection was organized through four different instruments: i) semi-structured interviews with ROs managers; (ii) web survey applied to project coordinators from ROs (62 responses from 63 projects – or 98% of response rate); (iii) web survey with counterparts of projects in firms (44 responses from 63 projects – or 70% of response rate); and (iv) semi-structured interviews conducted by five experts specifically hired to technically evaluate a sample of 25 projects.

The idea was to gather, by applying the different instruments, qualified information to measure input, output and also behavioral additionality, considering the perceptions of internal actors – directly target by EMBRAPII, such as ROs managers, project coordinators from ROs and counterparts of projects in firms – as well as the perceptions of external actors (the experts).

Data collected from those four instruments were analyzed using descriptive and multivariate statistics. In addition they were compared to each other in order to identify if and to which extent perceptions from the different actors (RO's project coordinators, companies and experts) do converge. For the latter a "convergence indicator" was used.

EMBRAPII influence was measured employing the "redundant causality identifier" – RCI, proposed by Salles-Filho et al. (2010; 2011).[1] Results revealed that the EMBRAPII's model had an important weight in promoting both behavioral and output additionality. That suggests the EMBRAPII's model has accomplished most of its initial intends.

Concerning project outputs, evaluation results show that expected technological results such as new products, processes and methodologies, were achieved in the majority of projects – although ROs were more optimistic about this issue than companies when the answers form both sides were compared. Those results were predominantly perceived as new to the country and in not-few-cases as new to the world. They were considered satisfactory by firms, taking part of their broad strategic plans. Intellectual property rights were generated in more than 50% of the projects. Moreover, projects contributed to the creation of new research areas or the consolidation of existing ones both in ROs and firms.

Experts confirmed that projects were pre-competitive in their design and execution, as foreseen in the EMBRAPII's model. Some firms were already able to use project's results in their internal processes or to commercialize these results, meaning that companies reported innovations. Impacts from these innovations are expected primarily in terms of added value and quality improvement, but also in revenues and market share. One important finding refers to the low importance of impacts in creating new business models and expanding exports. Experts were more pessimistic about impacts than the firms.

Beyond projects' output and outcomes ROs improved their research and innovation management processes, such as the ones related to prospecting opportunities and partners, negotiating and contracting projects, managing projects and raising financial resources. A Multiple Correspondence Analysis followed by a Cluster analysis showed evidences that organizational traits may have influenced the differences in outputs and outcomes. Particularly, the legal and managerial models in which the three RO's are based on seemed to have much to do with this. Some behavioral changes were found also at firms, although the influence of EMBRAPII's model has been perceived as less evident.

There are differences in perceptions of ROs, firms and experts concerning project's outputs and outcomes and behavioral changes. They occurred mainly about how projects were motivated, allocation of human and material resources from firms in project development and, as pointed out before, about the achievement of expected technological results, innovation and their impacts.

Although evaluation show more success in the achievement of technological results than in innovation itself, this seems to be a matter of timing, since firms showed satisfaction with almost all projects executed in the Pilot Phase. Results are consistent with evidence from other studies evaluating similar policies (Bienkowska et al., 2010; Cunningham and Gök, 2012; Marzucchi et al., 2015; Martin et al., 2015). Nevertheless, positive effects on the innovation efforts of Brazilian industry depend on long-term, stable commitment of government funding and support, including a reinforcement and expansion of EMBRAPII's model in years to come.

The preliminary conclusion about EMBRAPII model, based on evidences of evaluation of its pilot phase, is that the model is pretty effective in promoting linkages between ROs and firms towards R&D and innovation. Three main reasons can be raised to explain the success of the model: a) it induces contracts among ROs and companies giving them freedom to negotiate objectives and conditions and requiring mutual involvement in terms of financial support and managerial assistance; b) it facilitates the financial and operational conditions to execute projects leaving project's governance to the parties; c) it induces ROs to develop professional skills in R&D and innovation planning and management.

Notwithstanding, the evaluation showed gaps in the governance between ROs and companies, particularly related to asymmetries of information on project's outputs and outcomes. Depending on the type of indicators (patent filed for instance), parties did not converge in reporting the same outputs and outcomes.

Another interesting finding refers to the differences in the ROs performances. Under the same policy conditions, and during the same period, the three ROs performed in slightly different ways, particularly in terms of number of projects contracted – 30, 20 and 13 projects contracted by CIMATEC, IPT and INT, respectively -, and in flexibility to adapt their internal procedures and capabilities to the policy requirements – easier to CIMATEC, not difficult to IPT and more difficult to INT. The main hypothesis behind these differences refers to the institutional and managerial models and cultures upon which those organizations are built.

Finally, the evaluation can be considered well succeeded in creating qualified information to understand and inform policy-making process. The four instruments for collecting data seemed to be important to capture diverse perceptions about the same object and, afterwards, to understand the process of interaction and decision-making among actors, including companies, ROs and EMBRAPII itself as the commissioner of the policy.

1 The RCI is a sort of "what if" question and is useful in situations was there is no possible control group, which is the case of the EMBRAPII's Pilot.

REFERENCES

BIENKOWSKA, D., K. LARSEN, K., S. SÖRLIN (2010): "Public-private innovation: mediating roles and ICT niches of industrial research institutes," Innovation: Management, Policy & Practice, vol. 12, n. 2, pp. 206-216.

CUNNINGHAM, P., A. GÖK, A. (2012): "Impact of Innovation Policy Schemes for Collaboration". Part of the Compendium of Evidence on the Effectiveness of Innovation Policy Intervention. London: NESTA.

EDLER, J., M. BERGER, M. DINGES, A. GÖK (2012): "The practice of evaluation in innovation policy in Europe," Research Evaluation, vol. 21, n. 3, pp. 167–182.

HOCHSTETLER, K., A.P. MONTERO (2013): "The Renewed Developmental State: The National Development Bank and the Brazil Model," The Journal of Development Studies, vol. 49, n. 11, pp. 1484-1499.

KANNEBLEY JR, **S., G. PORTO** (2012): "Incentivos fiscais à pesquisa, desenvolvimento e inovação no Brasil: uma avaliação das políticas recentes", Discussion Papers, 236. Washington: Interamerican Development Bank.

LINK, A.N., N.S. VONORTAS (Ed.) (2013): "Handbook on the Theory and Practice of Program Evaluation". Cheltenham: Edward Elgar.

MARTIN, S., M. PAHOR, M. JAKLIC (2015): "The structure of policy-induced innovation networks in Slovenia," European Journal of Innovation Management, vol. 18, n. 4, pp. 428 – 450.

MARZUCCHI, A, D. ANTONIOLI, S. MONTRESOR (2015): "Industry–research co-operation within and across regional boundaries: what does innovation policy add?" Papers in Regional Science, vol. 94, n. 3, pp. 499-524.

PACHECO, C.A., S. CORDER (2010): "Mapeamento institucional e de medidas de política com impacto sobre a inovação produtiva e a diversificação das exportações," Documentos de Proyectos. Comissão Econômica para a América Latina e o Caribe (CEPAL).

PATTON, M.Q. (2012). "Essentials of Utilization-Focused Evaluation", Sage Publications.

SALLES-FILHO, S., F.D. AVILA, J. SEPULVEDA, F.A.B. COLUGNATI (2010): "Multidimensional assessment of technology and innovation programs: the impact evaluation of INCAGRO-Peru," Research Evaluation, vol. 19, n. 5, pp. 361-372.

SALLES-FILHO, S., M.B. BONACELLI, A.M. CARNEIRO, P.F.D. CASTRO, F.O. SANTOS. (2011): "Evaluation of ST&I programs: a methodological approach to the Brazilian Small Business Program and some comparisons with the SBIR program," Research Evaluation, vol. 20, n. 2, pp. 157-169.

SALLES-FILHO, S., G. N. STEFANUTO, C. V. MATTOS, C. ZEITOUM, F. R. CAMPOS (2012): "Avaliação de impactos da Lei de Informática: uma análise da política industrial e de incentivo à inovação no setor de TICs brasileiro," Revista Brasileira de Inovação, vol. 11, n.esp, pp. 191-218.

SUZIGAN, W., E. ALBUQUERQUE, R. GARCIA, M. RAPINI (2009): "university and industry linkages in Brazil: some preliminary and descriptive results," Seoul Journal of Economics, vol. 22, n. 4, pp. 591-611.

WHITLEY, R., J. GLASER (Eds.) (2007): "The Changing Governance of the Sciences: the advent of research evaluation systems". Springer.

AUTHORS

SERGIO SALLES-FILHO
Department of Science and Technology Policy, Institute of Geosciences
University of Campinas
R. João Pandiá Calógeras, 51, Campinas, SP 13083-870
Brazil
Tel: +55 19 3521-1280 (Corresponding author)

ADRIANA BIN
School of Applied Sciences, University of Campinas, R. Pedro Zaccaria, 1300, Limeira, SP 13484-350
Brazil

RAFAELA M. ANDRADE
Department of Science and Technology Policy, Institute of Geosciences
University of Campinas
R. João Pandiá Calógeras, 51, Campinas, SP 13083-870
Brazil

PAULA F. D. CASTRO
Department of Science and Technology Policy, Institute of Geosciences
University of Campinas
R. João Pandiá Calógeras, 51, Campinas, SP 13083-870
Brazil

FERNANDO A. B. COLUGNATI
Medical School, Federal University of Juiz de Fora, Rua José Lourenço Kelmer, 1300, Juiz de Fora, MG 36036-330
Brazil

NICHOLAS VONORTAS
Elliott School of International Affairs, The George Washington University, 1957 E Street, NW
Washington, DC 20052
USA

EVALUATING 'EXCELLENCE' IN THE ERC PEER REVIEW PROCESS

HELENE SCHIFFBAENKER AND PETER VAN DEN BESSELAAR

INTRODUCTION

The concept of scientific excellence is of increasing relevance for distributing research funds. The evaluation of excellence has become a core challenge in selection procedures like peer review processes. The European Research Council (ERC) is a highly competitive research funding organization operating within Horizon 2020. It aims to fund the highest quality and frontier research. The assessment of ERC grant applications is based solely on the criterion of scientific excellence. ERC grants are prestigious funds with high career impact. The construction and practice of excellence in ERC peer review panels is crucial when selecting the best quality science and scientists. The comparison of success rates for men and women in ERC peer review process raises the suspicion that assessment procedures are gender biased, as women are less successful in all ERC funding schemes. How may this be related to the construction of excellence and the peer review process? Gender biased outcomes underline the need to evaluate this selection mechanism.

While excellence is related to the meritocratic understanding of science - that success is based on individual performance and merit only - recent research has demonstrated that excellence is socially constructed (O'Connor and O'Hagan 2015, Rees 2011, Lamont 2009, Brouns and Addis 2004). The specific meaning and definition of excellence varies between scientific disciplines and fields, cultural and geographic contexts and between individual preferences and gender stereotypes (Heilman et al. 2015). Thus the concept of excellence is not gender neutral: As it reflects the norms of a masculine science system embodied in the stereotypical image of the ideal male scientist (Bailyn 2003, Benschop and Brouns 2003, Acker 1992). Gender bias is only one of various forms of bias when excellence is assessed in the peer review process (Langfeldt 2004).

Operationalizing excellence is a challenge for research funding organizations. In order to increase transparency, formal criteria have been defined by the ERC to measure excellence and to guide panelist in the assessment process. However, research suggests that criteria and related indicators typically used to describe excellence are gendered (Rees 2011, van den Brink and Benschop 2012). Past performance indicators such as number of publications have widely been criticized for not being gender neutral as the lower amount of time women are able to devote to research activities due to care responsibilities and unpaid work lowers their productivityand that the hierarchical position of researchers explains their productivity: as women are lower positioned in the science field they publish less (van den Brink and Benschop 2012, Aksnes et al. 2011).

RESEARCH APPROACH

This poses two major challenges for research funding organizations: the first challenge is related to the definition of scientific excellence and the question how gender bias is already inscribed into its definition and into the specific indicators to measure/assess excellence. Indicators strongly affect what is perceived as excellence and thereby determine who will be perceived as an excellent researcher. The other challenge comes into play when panellists or reviewers deploy formalized criteria and indicators to specific proposals. This raises the question how criteria are put in practice by panels, panellists and reviewers in specific contexts and how the peer review process is organised. This raises the question whether criteria are equally applied to male and female applicants or if so called gendered practices (Martin 2003) can be observed which means that criteria are applied differently to women and men.

Therefore we will investigate in this paper how criteria of scientific excellence and their application by panellists are gendered. In a first step of analysis we focus on how criteria for measuring scientific excellence are operationalized and deployed in practice. This will enable us to show in detail the shortcomings of peer review processes in applying objective assessments of scientific excellence. In a second step we analyse how these practices give space to gender practices and to a gender bias producing unequal evaluation outcomes (success rates). The paper is based on a study[1] commissioned by the ERC (2014-2016) to investigate the reasons for lower success rates of women in the prestigious ERC Starting Grant (StG).

DATA AND METHODS

We study the 2014 Starting Grant, involving 3200 applicants and 350 panelists. We conducted 32 semi-structured interviews with panelists, on a variety of aspects of the selection process – among others the way of deploying criteria for assessing scientific excellence. The interviews were transcribed, coded and analyzed. Furthermore, we compiled evaluation and past performance data to find out if gender differences in success rates could be explained by past performance differences and if scoring is consistent in the course of the peer review process.

1 This paper is an outcome of the project 'gendERC – gendered dimensions in ERC grant selection' (04/2014 – 02/2016), conducted for European Research Council (ERC) by Helene Schiffbaenker and Florian Holzinger (JOANNEUM RESEARCH) in close cooperation with Peter van den Besselaar and Claartje Vinkenburg (VU Amsterdam); Lucia Polo and Ezekiela Arrizabalaga (tecnalia). More information can be found at www.joanneum.at/policies.

FINDINGS

The interviewed panel members found most criteria and indicators related to excellence rather vague and difficult to measure, making comparison problematic. This gives space to individual interpretations. Further, criteria in general are deployed unsystematically in the grant selection process. This again gives space to gendered forms of deploying criteria that reflect unconscious gender stereotypes in the supposedly standardized selection procedures.

CONCLUSIONS

We conclude that for reducing gender bias following strategies should be developed: minimizing the subjectivity of assessments, defining assessment criteria more detailed, stronger standardization of assessment processes and raising awareness of panellists on how gender bias emerges in assessment practices. We will present suggestions how criteria could be operationalized and applied in a more effective way to arrive at gender-unbiased practicing of excellence and to improve transparency and the peer review process in general.

REFERENCES

Acker, J. (1992): "From sex roles to gendered institutions", Contemporary Sociology, vol. 21(5), pp. 565–569.

Aksnes, D. W., K. Rorstad, F. Piro, and G. Sivertsen (2011): "Are female researchers less cited? A large-scale study of Norwegian scientists", Journal of the American Society for Information Science and Technology, vol. 62(4), pp. 628–636.

Bailyn, L. (2003): "Academic careers and gender equity: Lessons learned from MIT" Gender, Work & Organization, vol. 10(2, pp. 137–153.

Benschop, Y. and M. Brouns (2003): "Crumbling ivory towers. Academic organizing and its gender effects", Gender, Work & Organization, vol. 10(2), pp. 194–212.

Brouns, M. & Addis, E. (2004): Synthesis report on the workshop. European Commission (2004): Gender and excellence in the making.

EC European Commission (2013): ERC Work Programme 2014.
Heilman, M. E., F. Manzi and S. Braun (2015): "Presumed incompetent: perceived lack of fit and gender bias in recruitment and selection", in Broadbridge, A. M. and S. L. Fielden (Eds.): Handbook of gendered careers in management. Getting in, getting on, getting out, ,Cheltenham, Edward Elgar, 2015.
Lamont, M. (2009): "How professors think: inside the curious world of academic judgment", Harvard College.

Langfeldt, L. (2004): "Expert panels evaluating research: decision-making and sources of bias", Research Evaluation, vol. 13(1), pp. 51–62.

Martin, P. Y. (2003) Said and Done" Versus "Saying and Doing". Gendering Practices, Practicing Gender at Work, in: Gender and Society, 17(3) 342–366.

O'Connor, P. and C. O' Hagan (2015): "Excellence in university academic staff evaluation: a problematic reality?", Studies in Higher Education.

Rees, T. (2011): "he gendered construction of scientific excellence", Interdisciplinary Science Reviews, vol. 36(2), pp. 133–45.

Van den Brink, M. and Y. Benschop (2012): Gender practices in the construction of academic excellence: Sheep with five legs. Organization, vol. 19(4) , pp. 507–524.

AUTHORS

HELENE SCHIFFBAENKER
JOANNEUM RESEARCH Forschungsgesellschaft mbH,
POLICIES – Institute for Economic and Innovation Research,
Sensengasse 1 1090 Vienna,
Austria
+43 1 58175202826
helene.schiffbaenker@joanneum.at

PETER VAN DEN BESSELAAR
VU University Amsterdam Department of Organization, Sciences & Network Institute
De Boelelaan 1105, 1081 HV Amsterdam
The Netherlands
+31 20 59 85376
p.a.a.vanden.besselaar@vu.nl

MEASURING THE IMPACT OF LARGE SCALE EUROPEAN FTI INTERVENTIONS: THE EU FLAGSHIP PROJECTS

MANFRED SPIESBERGER, FLORIAN KNECHT, KATHARINA BÜSEL, INGRID CLEMENT, STEFANIE KONZETT-SMOLINER AND HELMUT GASSLER

INTRODUCTION

In this article we will investigate the question of how to measure the impact of the EU Flagship projects. Flagships are long-term, very large scale research initiatives aiming to solve an ambitious challenge such as understanding the human brain or exploiting the potential of graphene, the newly discovered revolutionary material (European Commission, 2014). In October 2013 the first two EU Flagships, the Human Brain Project (HBP, https://www.humanbrainproject.eu/) and Graphene (http://graphene-flagship.eu/) started operation. These long-term initiatives are planned for a run-time of about 10 years and should receive and generate an investment of Ð 1 billion each over the run-time. They bring together excellent European research groups across various disciplines working on these topics. Flagships aim at transformational impacts on science and technology, delivering a key competitive advantage for European industry and substantial benefits for society.

The impact measurement approach presented in this article is based on work performed in the EU Horizon 2020 funded Coordination and Support Action TAIPI - Tools and Actions for Impact Assessment and Policy makers Information (https://taipi.eu/). TAIPI started operation in January 2015 and supports the Flagships in their impact evaluation and in communicating the results of the impact evaluation.

METHODOLOGY

In a first step in the TAIPI project we have developed an assessment frame for measuring the impact. This was based on a comprehensive literature review and on interviews with 20 leading evaluation experts working mainly on programme evaluation. The interviewees were from research funding organisations such as the Austrian Science Fund and the Finnish TEKES, from the European Commission, OECD, Fraunhofer, and other relevant national and international organisations.

In a second step we have specified the indicators to be used for the Flagship impact assessment. A workshop was held among the TAIPI partners for developing the indicators. The indicators were then classified into easy and difficult to collect, and then cross checked with the Flagship management.

MEASURING EU FLAGSHIP IMPACT

In a third step, which is currently in the implementation phase, we are collecting the data for measurement of the specified indicators. This involves provision of quantitative data by Flagship management, conducting surveys among Principal Investigators involved in the Flagships, and conducting interviews for collecting qualitative information (e.g. for case studies).

MEASURING THE IMPACT

Interviews with the evaluation experts and literature review guided us in specifying the assessment frame. A key message was to use mixed-methods when assessing the impact of the flagships. Case studies, expert interviews, interviews with funded researchers and focus groups are needed for understanding and interpreting quantitative indicators. Another important lesson learned was that data and indicators should concern formative aspects and inputs, publication output, commercial valorisation, international networking and interdisciplinary cooperation. For a stronger focus on impact, it will be necessary to look at the "hot papers" and "highly cited papers", personnel exchanges, spin-offs, patents, etc. A focus should be on simple indicators, while composite indicators should be avoided in the case of the Flagships, as they have only limited added value for this impact evaluation. Bibliometric and patent analysis should complement indicators.

On the basis of this input and by studying basic documents of the Flagships and the European Commission, we suggested an assessment frame. We classified the expected impacts of the Flagships into 6 categories. Single indicators were specified per each impact dimension.

Structural impact involves indicators such as the partnership of the Flagships and scientific disciplines represented. Cooperation and collaboration impact involving indicators such as international partners (beyond the EU) cooperating in the Flagship and interdisciplinary research. This will be complemented by a Social Network Analysis to analyse cooperation patterns within the Flagships and on the other hand, to find out whether new networks were established, due to the Flagships.

Scientific impact is the most evident to analyse. Excellence of research (e.g. to look at hot papers and highly cited papers): Check the highly visible parts (in terms of citations) of the Flagships' output rather than the entire and average output.

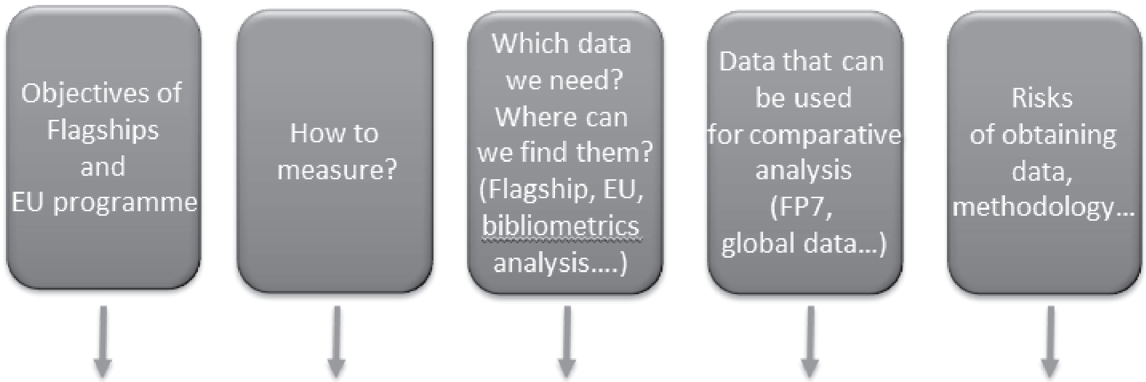
Economic impact "is the engine of the other components". Although this impact is very important it is difficult to link research activities to job creation or productivity increase. There has to be a focus on measurable data like partnerships with industries or patent analysis.

Social impact covers many dimensions, including progress for health, acceptability of new products, education by research, policy diffusion and responsible research and innovation (RRI). Typical RRI components are: Gender, Public Engagement, Ethics, Science education, Governance and Open access. "Science communication": are they taken up in e.g. daily press and public media? Ethics: How many ethic commissions are implemented in the Flagships? How many ethical audits are organised?

Environmental impact is relevant mostly for the Graphene Flagship, because it is expected to generate environmentally friendly technologies.

Measuring EU Flagship impact

Figure 1: Assessment Frame for EU Flagships' impact



Expected impact	Indicators	Data and sources	Comparison elements	Risks
Structural				
Cooperation and collaboration				
Scientific				
Economic				
Social				
Environmental				

CHALLENGES

In the assessment of the Flagships' impact, we have to consider several challenges:

Flagships are collecting themselves quantitative Knowledge and Performance Indicators (KPIs). These include number of publications, number of citations, number of educational courses and attendees, and others. In the assessment by TAIPI, we have to observe that we have a good complementarity to existing KPIs and to avoid duplication of effort. In particular are lacking in the KPIs qualitative measurement and social issues, which have therefore to be dealt with in the TAIPI assessment.

The management work to implement the Flagships and to coordinate partnerships of about 110 partner organisations in HBP and about 170 in Graphene is significant. This has left the impact assessment a bit at the side-lines of management. Management teams have few time to dedicate to this work, and it needs some efforts at persuasion to organise data collection.

The Flagships have at this point only a short run-time of two and a half years. This inhibits making assessments on longer term impact. For example scientific publications, which are a key indicator for the Flagships and the European Commission, have a certain time-lag until they are published. Publications are taking-up momentum in the Flagships only now, after a certain period of support.

Significant differences between the two existing Flagships need to be observed. While HBP is a basic research oriented project developing towards a research infrastructure, Graphene is more applied research oriented with a stronger focus towards immediate economic benefit. This has repercussions on indicators. For example spin-offs are not relevant yet for HBP as an indicator, while they are for Graphene.

MEASURING EU FLAGSHIP IMPACT

The Flagships are rather unique initiatives from the point of view of their size and ambition. It is therefore difficult to find suitable initiatives for comparison. Possible comparison could be the EU's Framework Programme 7 for Research and Development, or similar national initiatives such as the US brain initiative.

REFERENCES

European Commission (2014) COMMISSION STAFF WORKING DOCUMENT: FET Flagships: A novel partnering approach to address grand scientific challenges and to boost innovation in Europe, SWD(2014) 283 final

AUTHORS

MANFRED SPIESBERGER
Centre for Social Innovation (ZSI)
Linke Wienzeile 246, 1150 Vienna
Austria
Mobile Phone: +43-650-6812122
spiesberger@zsi.at

FLORIAN KNECHT
Erdyn
Paris
France

KATHARINA BÜSEL
Centre for Social Innovation (ZSI), Vienna
Austria

INGRID CLEMENT
Erdyn
Paris
France

STEFANIE KONZETT-SMOLINER
Centre for Social Innovation (ZSI), Vienna
Austria

HELMUT GASSLER
Centre for Social Innovation (ZSI), Vienna
Austria

PIERRE-BENOÎT JOLY
Université Paris-Est
LISIS 77454 Marne-la-Vallée,
France

CHALLENGES OF EVALUATING COMPLEX EUROPEAN POLICY INITIATIVES: CASE OF EUROPEAN RESEARCH AREA

INGA ULNICANE

The aim of this paper is to discuss challenges of evaluating complex European research policy initiatives such as the European Research Area (ERA). It reflects on lessons learned from a recent study on the ERA initiative commissioned by the European Parliamentary Research Service (EPRS) and carried out by the author. The EPRS requested to identify gaps and barriers in the ERA initiative and to suggest recommendations for future policy activities including possible legislation. This paper aims to go beyond the study submitted to (November 2015) and presented at (April 2016) the European Parliament by drawing on a number of challenges encountered during it that can be relevant for future research and policy discussions about the ERA initiative and European research policy.

During the last 15 years there have been many studies, projects and evaluations of the ERA initiative, including the Commission's own ERA monitoring exercise in recent years. A relatively small study commissioned by the EPRS was interesting as instead of going into details of specific ERA priorities it allowed to take a broader view on the ERA initiative and possible policy actions.

This extended abstract will firstly briefly introduce the ERA initiative and gaps and barriers identified in it and then discuss challenges encountered and lessons learned during the above mentioned study.

THE ERA INITIATIVE: GAPS AND BARRIERS

The ERA is an ambitious and complex policy initiative that the European Commission launched in 2000 aiming at free circulation of researchers, scientific knowledge, and technology (Ulnicane, 2015). During 16 years its aims and governance has evolved, and today it is structured around the six ERA priorities of effective national research systems, transnational cooperation, an open labor market for researchers, gender equality, optimal circulation, and international cooperation. Each of these six policies priorities further involves many activities such as joint programming, the charter and the code for recruiting researchers, and a scheme for portability of supplementary pensions of internationally mobile researchers. This combination of a number of priorities and many activities makes analysis and evaluation of the ERA initiative a highly complex task.

To identify gaps and barriers in the ERA initiative, a relatively small study (prepared July-October 2015), combined evidence from multiple sources including review of EU policy documents on the ERA initiative, previous studies and evaluations, academic research, and carried out 16 interviews with relevant policy makers and stakeholders. This was supplemented by secondary analysis of approximately 100 interviews with researchers, policy makers, and stakeholders that the author has carried out within her previous research projects. Upon the request of the EPRS, the study focused on the first three ERA priorities – first, effective national research systems, second, transnational cooperation, and third, an open labor market of researchers – with a special focus on the third one.

This led to identification and discussion of 14 gaps and barriers in the ERA initiative, a number of which have also been identified in earlier studies and evaluations. Gaps and barriers that apply to the overall initiative are first, insufficient coordination with other policies and initiatives (e.g. European Higher Education Area, Structural funds, innovation policy), and second, limited range of interests represented in the ERA stakeholder platform that in 2015 included only five organizations of research performers and funders from a very broad range of relevant stakeholders including also grass-root organizations of scientists and business interests. Specific gaps and barriers regarding the first ERA priority on more effective national research systems are uneven progress across member states and narrow focus on project-based funding as an indicator for effective national research systems. For the second ERA priority on jointly addressing grand challenges, two shortcomings were identified, namely, lack of output evaluation of jointly addressing grand challenges as well as lack of support for bottom-up trans-national research collaboration.

As requested, the main focus of the study was on the third ERA priority on open labor market for researchers, aiming to ensure the removal of barriers to researcher mobility (within and from outside of the EU as well as inter-sectoral), doctoral training and attractive careers. Here eight gaps and barriers were identified: unidirectional flows of researchers from South/East to North/West of Europe; lack of open, transparent and merit-based recruitment; low participation in EU initiatives establishing open labor market for researchers (the Charter and Code for recruitment, RESAVER pension scheme); limited portability of and access to national research grants; limited dual career opportunities; language barriers; unclear demand for researchers; and job insecurity. Policy recommendations were reviewed to address these gaps.

CHALLENGES FOR EVALUATING THE ERA INITIATIVE

What are the lessons learned and challenges identified from this commissioned study that can be useful for future research and evaluations? Four challenges can be outlined here: first, developing meaningful links between European research policies and practices; second, broadening and diversifying evidence base on ERA aims on international mobility, collaboration and competition; third, addressing interdependencies of the ERA priorities of efficient national research systems, collaboration, and open labor market for researchers; and fourth, developing recommendations that move forward voluntary coordination vs. legislation tension. These four challenges will be briefly discussed.

First challenge of developing meaningful links between European research policies and practices is related to a tension that research practices are international while policies are largely national (Nedeva, 2013). The activities of the ERA initiative largely take place at the policy level but also aim to facilitate international research practices. Challenge is to understand and evaluate the role of EU policies in the field of research where practices of collaboration and mobility have been international for centuries and today are increasingly so due to many scientific and other reasons. While the EU policy discourse typically describes EU activities as 'added value', in the field of research they might rather be seen as policy attempt to 'catch-up' with long established practices in the scientific community and intrinsic needs of knowledge production. While a lot of research have been done on both – micro-level research practices and macro-level policies – there might be further opportunities in establishing meaningful links between both that could also help in better conceptualization of complex multi-level initiatives such as the ERA.

Second challenge draws attention to the need to broaden and diversify evidence base on the ERA aims on international mobility, collaboration and competition. The ERA policies depict collaboration, competition, mobility, peer-review, project-based funding and other aims and indictors as highly beneficial that should be further promoted. While in academic and policy research there is a lot of evidence of positive benefits from these process, there are also findings demonstrating that these aims and indicators can have downsides, be counter-productive and lead to inefficiencies, for example in cases of forced mobility or collaboration, too high dependence on and competition for project-based funding (Stephan 2012). Challenge here is to incorporate these positive and negative aspects in evaluations and to come up with more nuanced policy recommendations when and under what conditions collaboration, competition and mobility is beneficial.

Third challenge for analyzing the ERA initiative is to address interdependencies among different ERA priorities such as efficient national research systems, collaboration, and open labor market for researchers. As each ERA priority include very specific aims and activities, governance and assessment of the initiative tends to be 'compartmentalized' with experts in specific areas - mobility, joint programming, gender – governing and evaluating each of them separately. However, the ERA priorities are also tightly connected and interdependent: only among effective national research systems (ERA priority 1) it is possible to have beneficial international mobility (ERA priority 3) and collaboration (ERA priority 2). Looking on interactions between different priorities reveals deeper long-standing problems showing for example that uneven progress of reforming national research systems (ERA Priority 1) leads to unidirectional flows of researchers from South/East to North/West of Europe (ERA Priority 3). Thus, there is 'added value' in finding a more 'holistic' approach to governing and evaluating the ERA initiative that takes interdependencies of the ERA aims and priorities into account.

Fourth challenge addresses the task of many evaluations, namely, developing policy recommendations. One of the long-standing divisions (among different EU institutions and stakeholder groups) regarding the ERA initiative has been between those who support the ERA development by voluntary coordination (as it has largely been so far) and those who argue for the need to introduce legislation. Proponents of voluntary coordination typically argue that it is better in accommodating national diversity and specificity of research, while supporters of legal measures present these as more efficient way to promote progress in the ERA implementation which they often see as insufficient. One challenge for analyzing potential of legal route is that its proponents usually have been rather general without specifying what kind of legal measures might help to implement the ERA priorities. One of rare assessments of the legal options for the ERA (Pilniok 2014) after reviewing diverse options (hard and soft law, framework directive or sector specific regulations) advises caution and reminds that "the role of (binding) law should not be overestimated as a steering mechanism for the research system"; if political decision for a legislative option is made, Pilniok (2014) suggests to focus on removing barriers to mobility (recruitment, access to and portability of grants, pensions) but points out that these still might encounter difficulties due to differences in attractiveness of national systems. One way to move forward the tension between voluntary vs. legislative options for the ERA initiative is to look for other instruments. Recently EU research funding mechanisms (Horizon 2020 and Structural funds) have introduced some conditions for receiving funding (on recruitment and national reforms); it is a task of future evaluations to assess if such conditionality might be an option for facilitating implementation of the ERA.

It is relevant to discuss these challenges because the ERA initiative is currently on the agenda of European institutions including national ERA roadmaps and new ERA monitoring exercise to be undertaken in 2016. Moreover, a number of issues mentioned in this abstract relates to broader issues in evaluation of European research policies.

REFERENCES

Nedeva, M. (2013): "Between the global and the national: Organising European science," Research Policy, vol.42, pp. 220-30.

Pilniok, A. (2014): Translating ERA into legislation? How to ensure the completion of ERA. European Commission.

Stephan, P. (2012): How economics shapes science. Cambridge, MA: Harvard University Press.

Ulnicane, I. (2015): "Broadening Aims and Building Support in Science, Technology and Innovation Policy: The Case of the European Research Area," Journal of Contemporary European Research, vol.11, pp. 31-49.

AUTHOR

DR. INGA ULNICANE
Institute for European Integration Research EIF
University of Vienna
Strohgasse 45/DG 1030 Vienna Austria
Tel: +431427722414
inga.ulnicane@univie.ac.at

STAKEHOLDER INVOLVEMENT IN EVALUATION PROCESSES

FRITZ, M.M.C

In the last decades, the stakeholder theory has been widely applied in a variety of research fields like social sciences or business management. Recently, researchers and practitioners have been integrating the concept in policy-making, especially for issues related to environmental protection (Freeman et al., 2010, p.182). It is believed that stakeholder analysis techniques can support the development of successful policies by understanding the different wishes and engaging with multiple stakeholders at multiple levels (local, regional, national and international).

However, to engage with stakeholders it is necessary to identify them and know what their present and future roles and their direct or indirect influences may be. This essential step in stakeholder analysis processes is often undervalued and conducted with bias (Reed et al., 2009), if described at all, leading to the omission of key stakeholders and policy failures or unexpected outcomes. Also, stakeholder analysis tools come from fields that may not be adapted for policy-making since they often place a firm at the centre of the analysis. Such approaches are even becoming obsolete in certain sectors like energy where energy users and homeowners can also produce and store their own energy which converts them into energy suppliers (Fritz et al., 2016a). Thus stakeholder identification is crucial and the way stakeholders can be identified shall also be guided in a policy-making context to contribute to positive evaluation and success of policies.

This paper hence suggests adapting existing tools for stakeholder identification to the policy-making context. This is based on the development, testing and assessment of the Supply Chain-Oriented Procedure for Identifying Stakeholders (SCOPIS) that places the product at the centre of the analysis instead of a firm and uses well-known scientific methods like literature reviews or interviews that allow replicability of the approach from local to global policy-making and engagement with stakeholders (Fritz et al., 2016b). This procedure gathers well-known scientific methods such as literature review, interviews, and questionnaires which allows replication of the approach. The procedure also integrates the requirements from scientific papers that highlighted the need for iterative processes, visualisation tools, and the consideration of time and context to reduce bias and omission risks (e.g., Reed et al. 2009; Bryson, 2004; Bourne and Walker, 2006) and support the understanding of multi-level and multi-temporal issues (e.g., Salado and Nilchiani, 2013; Achterkamp and Vos, 2007). SCOPIS has been fully tested and assessed in the case of 1) mercury use in Artisanal and Small-scale Gold Mining (ASGM) and to a certain extent in the case of 2) an Austrian city located in Styria, which is in transition towards sustainable business models for energy supply. These two cases corroborate the application of the process at a global and regional level.

The use of this procedure to identify stakeholders related to a product (e.g., mercury contamination) or service (e.g., new energy-related services) present several advantages compared to traditional stakeholder analysis approaches. First, placing a product or service at the centre of the analysis and identifying the stakeholders involved in producing, supplying and using this product or service enables to set a strong basis to stimulate discussion in focus groups or experts' interviews with the support of a visualization tool (e.g., diagram). Second, placing a product or service at the centre of the analysis is also a way to avoid bias and conduct research more ethically since one company alone can often not be held responsible for a specific problem (Bryson, 2004). Third, on the basis of the literature and assessment of the validity of the results by experts, the iterative process of engaging experts and other stakeholders (e.g., government officials, NGOs) enables to provide more objective results and identify stakeholders that may be omitted when only a certain group of experts is consulted once.

The use of the procedure to identify stakeholders can also present some difficulties, especially when engaging with companies. When engaging with experts to validate the stakeholders identified in the literature in Case 2, some resistance was observed at the beginning due to the placement of energy at the centre of the diagram instead of the client. But once the process was understood, an exchange took place which validated, enhanced and slightly contradicted the initial findings from the literature review and the bilateral meetings. When comparing the use of the procedure in Case 1 (global issue) and Case 2 (local issue), one may note that at a global level, the procedure requires more time and more work regarding the literature review, the identification and engagement of stakeholders. At a local level, it is easier to reach experts and engage them in a face-to-face or group meeting, which does not require going through all 8 steps suggested in the procedure. The procedure has been assessed by 9 different experts including policy-makers related to case 1. Further research would be required in order to assess the relevance of the approach at a local, regional and national level and in different policy contexts like the ones defined by the Millenium Development Goals (e.g., eradicate extreme poverty and hunger, reduce child mortality).

KEY REFERENCES

Achterkamp, M.C., Vos, J.F. J., 2007. Critically Identifying Stakeholders – Evaluating boundary critique as a vehicle for stakeholder identification. Systems Research and Behavioral Science, 24, 3-14. Doi:10.1002/sres.760.

Bourne, L., Walker, D., 2006. Using a visualising tool to study stakeholder influence–two Australian examples. Manag. Decis. 37, 1–29.

Bryson, J. M., 2004. What to do when stakeholders matter Stakeholder identification and analysis techniques. Public Management Review ISSN 1471-9037 print/ISSN 1471-9045 online. Doi:10.1080/14719030410001675722

Freeman, R.E., Harrison, J.S., Wicks, A.C., Parmar, B.L., de Colle, S., 2010. Stakeholder Theory – The State of the Art. Cambridge University Press: New York

Fritz, M.M.C., Rauter, R., Baumgartner, R.J., 2016a. Identifying stakeholders in a structured way - A basis for developing sustainable energy business models for cities, in: 22nd International Sustainable Development Research Society Conference (ISDRS 2016), Vol. 3, p.331, School of Science and Technology, Universidade Nova de Lisboa, Lisbon, Portugal, 13-15 July 2016.

Fritz, M.M.C., Rauter, R., Dentchev, N., 2016b (Submitted). A supply chain-oriented process to identify stakeholders.

Reed, M.S., Graves, A., Dandy, N., Posthumus, H., Hubacek, K., Morris, J., Prell, C., Quinn, C.H., Stringer, L.C., 2009. Who's in and why? A typology of stakeholder analysis methods for natural resource management. Journal of Environmental Management, 90, 1933–1949. Doi:10.1016/j.jenvman.2009.01.001

Salado, A., Nilchiani, R., 2013. Contextual- and Behavioral-Centric Stakeholder Identification. Procedia Comput. Sci. 16, 908–917. Doi:10.1016/j.procs.2013.01.095

AUTHOR

FRITZ, M.M.C
Karl Franzens University, Institute of Systems Sciences, Innovation and Sustainability Research
Merangasse 18/I, 8010 Graz
Austria
+43 (0)316 380 1525
morgane.fritz@uni-graz.at

EVALUATION FROM INSIDE? EVALUATING STRUCTURAL CHANGE PROCESSES TO PROMOTE GENDER EQUALITY

FLORIAN HOLZINGER, JUERGEN STREICHER AND HELENE SCHIFFBAENKER

INTRODUCTION

The promotion of gender equality has become a key priority for the European Research Area reform agenda. Member States, the European Commission (EC) and stakeholder organisations operating in the field of research, technology and innovation (RTI) are to promote and take actions to better achieve gender equality. One particular approach to promote gender equality, aiming to encourage institutional and cultural change, is by developing and implementing gender equality plans, a topic which has received strong support through calls in the European Research Framework Programmes.

These calls provide funding for research performing organisations (RPO) as well as research funding organisations (RFO) who commit themselves to develop gender equality plans and engage individual actors as well as the organisation in the modernisation of institutional practices. Proposals should aim to "increase the participation and career advancement of female researchers, improve working conditions of women and men as well as the integration of gender in curricula and research content" (EC 2012a, 28). Respective projects should address the norms, believes and values of scientific organizations (in RPOs and RFOs alike) entrenched in the formal and informal rules and procedures of these organizations (EC 2012b). Importantly, project consortia are required to include a detailed methodology and relevant steps to be used for monitoring and assessing the effectiveness and anticipated impacts of proposed actions as well as the institutional progress achieved, including, for instance, its impact on the number and situation of women scientists as well as on the integration of gender in research content (EC 2012b).

While gender equality has been on the RTI policy agenda of the European Union for more than a decade (EC 2010) and to a varied degree in its member states (EC 2008, 2014), a comprehensive and agreed upon methodology to measure outputs and impacts of structural and cultural change processes to promote gender equality is somewhat lacking. Past studies, such as the "She Figure" publications (see EC 2003, 2006, 2009, 2013, 2016), have directed substantial efforts to develop quantitative indicators and use different data sources to provide a longitudinal perspective on e.g. the (under)representation of women in RTI. However, they are less useful when measuring outputs and impacts of structural and cultural change projects in organisations.

As has been suggested by van den Brink and Benschop (2012; see also Stainback et al. 2015; Abrahamsson 2014), cultural and structural change processes are complex and may not lead immediately or directly to improved inequality gaps. They are context-bound and determined by many factors and mechanisms, and likely to be affected by setbacks or counter-productive feedback-loops, requiring a long term perspective to measure impacts of these processes on a quantitative level. However, EC funded structural change projects may last only for few years, making it difficult to follow up on these developments. Recent studies have focused great attention to this issue, suggesting adapted monitoring and evaluation approaches to assess the achievements of this kind of projects (see, for instance, Lipinsky and Schäfer 2014, Genova et al. 2014, Cacace et al. 2015). In this light, the ERA-NET project "GENDER-NET" has started to collect indicators to monitor implementation and to assess impacts of structural change processes on the organizational level (see Gender-Net forthcoming). Also, the US ADVANCE programme, role model for the EC structural change calls, has initiated further attempts to evaluate the outcomes of these projects (see Frehill 2006, Frehill and Kehoe 2006).

OBJECTIVE AND APPROACH

The objective of this paper is to propose a concept and indicators for evaluating the achievements of structural and cultural change projects that support gender equality based on gender equality plans and discusses initial results arising from two EU framework projects. These projects are "GARCIA" – Gendering the Academy and Research: Combating Career Instability and Asymmetries, which started in 2014 and will end in the beginning of 2017, and "GENERA" – Gender Equality Network in the European Research Area, which started in 2015 and will last until 2019.

Structural and cultural change processes take place in complex and dynamic systems (Garcia and Zazueta 2015, Byrne 2013). Outcomes may be generated in different ways, also drawing attention to the context in which a specific measure is applied, less so to the measure itself. Furthermore, especially cultural change processes are dealing with intangible and often informal aspects of organizations which make a straight forward way of measuring and quantifying difficult (Gherardi and Poggio 2001, Ely and Meyerson 2000). These aspects have to be taken into account by a practical evaluation concept.

In light of this, the proposed evaluation concept and indicators builds on work in evaluation research, especially theory based approaches (Blamey and Mackenzie 2007, Williams 2015), and develops a conceptual foundation based upon literature on structural and cultural change in research organization (see e.g. EC 2012b, Acker 2006, Benschop and Doorewaard 2012, Ely and Meyerson 2000; Pepin et al. 2014, Sappleton and Takruri-Rizk 2008, Abrahamsson 2014, Schein 1990). Instruments and indicators are developed that seek to assess the strengths, weaknesses, challenges and resistances of implementation processes as well as the outputs, outcomes and anticipated impacts. Indicators are qualitative and quantitative and are organised along main themes and aspects of gender equality identified in the literature. Focal point of the evaluation concept and indicators is assessing short to medium term achievements.

This paper uses mixed methods combining quantitative and qualitative approaches to collect and analyse data, including document analysis, interviews, survey results, secondary data analysis and self-reporting tools. The views and experiences of diverse actor groups, reaching from members of the implementation team and direct target groups (i.e., participants of implemented actions) to indirect target groups (i.e., people who are affected by measures targeted at a different target group) are considered. While the multiple perspectives helped developing a holistic view on implementation processes and what can be learned from them in terms of good implementation practices and impacts, they also constitute somewhat of a challenge. Preliminary results indicate that interpretation of project achievements may vary considerably amongst project partner, depending on e.g. goals set, progress of implementation and position within the project or the participating organization. The role of the evaluating team, which are integral part of the funded projects, may also shape implementation processes and is discussed in light of their dual roles as detached or "critical friend" and project partner.

EXPECTED RESULTS

This paper aims to advance the understanding of structural and cultural change projects that promote gender equality and their outcomes and impacts on the organisational level. It will present and discuss findings and lessons learned from the evaluation of outcomes and impacts of the GARCIA and GENERA project implementation processes. Resistances (e.g. lack of commitment, disregarding gender studies) and facilitators (e.g. change agent, mobilizing community discussions) related to the implementation processes are critically examined, and their relevance in collecting and analysing data are discussed. Results of output and outcome indicators obtained from the analysed projects are presented, and advantages as well as challenges of applying these indicators to monitoring and evaluating structural change processes are discussed.

Also, this study seeks to contribute to the emerging body of research on impacts of gender equality policies that aim to foster structural and cultural change. Findings may provide insights into opportunities and challenges of the respective implementation processes and into its complex and subtle achievements. This may also help facilitating better designed and implemented structural change processes in the future.

REFERENCES

Abrahamsson, L. (2014), "Gender and the modern organization, ten years after", Nordic Journal of Working Life Studies, Vol. 4 No. 4, pp. 109–136.

Acker, J. (2006), "Inequality Regimes Gender, Class, and Race in Organizations", Gender and Society, Vol. 20 No. 4, pp. 441–464.

Benschop, Y. a. Doorewaard, H. (2012), "Gender subtext revisited", Equal Div and Incl: An Int J, Vol. 31 No. 3, pp. 225–235.

Blamey, A. u. Mackenzie, M. (2007), "Theories of Change and Realistic Evaluation. Peas in a Pod or Apples and Oranges?", Evaluation, Vol. 13 No. 4, pp. 439–455.

Byrne, D. (2013), "Evaluating complex social interventions in a complex world", Evaluation, Vol. 19 No. 3, pp. 217–228.

Cacace, M., Balahur, D., Bleijenbergh, I.L., Falcinelli, D., Friedrich, M. a. Kalpazidou Schmidt, E. (2015), Structural Transformation to Achieve Gender Equality in Science: Guidelines, Rome.

European Commission (2003), She Figures 2003: Women and Science, Statistics and Indicators, Luxembourg.

European Commission (2006), She figures 2006: Women and Science: Statistics and Indicators, Luxembourg.

European Commission (2009), She Figures 2009: Statistics and indicators on gender equality in science, Luxembourg.

European Commission (2013), She Figures 2012: Gender in Research and Innovation, Statistics and Indicators, Brussels.

European Commission (2016), She Figures 2015, Luxembourg.

European Commission (2008), Benchmarking policy measures for gender equality in science, Luxembourg.

European Commission (Ed.) (2010), Stocktaking 10 years of "Women in Science" policy by the European Commission 1999-2009, Publications Office of the European Union, Luxembourg.

European Commission (2012a), Work Programme 2013, Capacities, Part 5, Science in Society, C(2012)9379

European Commission (2012b), Structural Change in research institutions: Enhancing excellence, gender equality and efficiency in research and innovation, Luxembourg

European Commission (2014), Gender Equality Policies in Public Research: Based on a survey among Members of the Helsinki Group on Gender in Research and Innovation, 2013, Luxembourg.

Ely, R.J. a. Meyerson, D.E. (2000), "Theories of Gender in Organizations. A New Approach to Organizational Analysis and Change", Research in Organizational Behavior, Vol. 22, pp. 103–151.

Frehill, L.M. (2006), Using Program Evaluation to Ensure the Success of Your ADVANCE Program.

Frehill, L.M. a. Kehoe, P. (2006), Toolkit for Reporting Progress Toward NSF-ADVANCE: Institutional Transformation Goals.

Garcia, J.R. u. Zazueta, A. (2015), "Going Beyond Mixed Methods to Mixed Approaches. A Systems Perspective for Asking the Right Questions", IDS Bulletin, Vol. 46 No. 1, pp. 30–43.

Gender-Net (forthcoming), Synthesis report on selected structural change initiatives and indicators for monitoring of state-of-play and progress

Genova, A., Micheli, B. de, Zucco, F., Grasso, C. a. Magri, B. (2014), Achieving gender balance at the topof scientific research: Guidelines and tools for institutional change, Rome.

Gherardi, S. u. Poggio, B. (2001), "Creating and recreating gender order in organizations", Journal of World Business, Vol. 36 No. 3, pp. 245–259.

Lipinsky, A. u. Schäfer, M. (2014), Detailed Evaluation Concept, Integer (D7.13)

Pépin, A., Collina, J., Pontois, M.T., Drew, E., Marshall, C., ŠidlauskienĐ, V., Jazdauskas, G., Lipinsky, A., Löther, A. a. Schäfer, M. (2014), Fostering Gender Equality in Research Institutions through Transformational-Gender Action Plans, 8th European Conference on Gender Equality in Higher Education, Vienna.

Sappleton, N. a. Takruri-Rizk, H. (2008), "The Gender Subtext of Science, Engineering, and Technology (SET) Organizations: A Review and Critique", Women's Studies: An inter-disciplinary journal, Vol. 37 No. 3, pp. 284–316.

Schein, E.H. (1990), "Organizational Culture", American Psychologist, Vol. 45 No. 2, pp. 109–119.

Stainback, K., Kleiner, S. u. Skaggs, S. (2015), "Women in power. Undoing or Redoing the Gendered Organization?", Gender & Society.

van den Brink, M. a. Benschop, Y. (2012), "Slaying the Seven-Headed Dragon. The Quest for Gender Change in Academia", Gender, Work & Organization, Vol. 19 No. 1, pp. 71–92.

Williams, B. (2015), "Prosaic or Profound? The Adoption of Systems Ideas by Impact Evaluation", IDS Bulletin, Vol. 46 No. 1, pp. 7–16.

AUTHORS

FLORIAN HOLZINGER
JOANNEUM RESERACH
Sensengasse 1, 1090 Vienna
Austria

JUERGEN STREICHER
JOANNEUM RESERACH
Sensengasse 1, 1090 Vienna
Austria

HELENE SCHIFFBAENKER
JOANNEUM RESERACH
Sensengasse 1, 1090 Vienna
Austria

DIVERSITY AMONG STAKEHOLDERS AND THE EVALUATION OF IMPACT AND RELEVANCE OF PUBLIC RESEARCH

A.A.M. PRINS AND J.B. SPAAPEN

RESEARCH ASSESSMENT IN A VARIEGATED CONTEXT

Much of the literature about research impact assessment stresses the importance of network approaches, suggesting a crucial role for interaction between researchers and stakeholders in innovation processes. Recent experiences in evaluation practices such as the British REF exercise show considerable diversity of such networks (Kings College and Digital Science 2015). Such diversity not only exists for academic research groups but also for public research organizations that address specific issues and problems related to governmental policy and/or professional interests.[1] Also, many of these public research organizations work in specific contexts such as national settings.

Although the diversity in tasks and missions and the diversity in contexts certainly poses a challenge for the assessment of the societal impact of research, it does not exclude a systematic approach to evaluation or assessment. In this paper we will address the issue of diversity among stakeholder networks in academic and public research organizations, and offer a systematic approach for analyzing these networks while discussing some of its limitations and the implications for impact evaluation procedures.

Networks of stakeholders focusing on societal issues are as a rule more diverse than traditional academic networks, but often also more volatile. They are characterized by a variety of academic stakeholders (various scientific and technical disciplines) and stakeholders from society, be it industry, government or society at large. Elsewhere, we have collaborated with many colleagues and stakeholders to analyze such networks and researched what the consequences could be for the evaluation (www.siampi.eu)[2]. The analysis in the SIAMPI project focused on the different types of interactions that take place between the stakeholders in such networks of research and innovation[3]: (1) Direct, in the sense of "personal" interactions that evolve around face-to-face encounters, or through phone, email or videoconferencing; (2) Indirect interactions through some kind of material "carrier": these include texts such as policy reports, protocols, books, music scores and questionnaires as well as artefacts such as websites, software, exhibitions, devices; and (3) Material interactions occur when potential stakeholders engage in a financial contribution, a contribution "in kind," or when facilities are shared. To research these interactions SIAMPI used a variety of methods. Among them were face-to-face interviews with academics and societal stakeholders and focus groups.

THE UPTAKE OF RESEARCH OUTPUT BY STAKEHOLDERS

In a number of studies among academic and public research organisations in the Netherlands some of the ideas developed in the SIAMPI project were tested, in particular using a methodological approach of the uptake by stakeholders of research output. This method, contextual response analysis (CRA), traces the uptake of written output to society and to chart the stakeholder context. The involved research organizations encompass a wide range of research fields and policy tasks, ranging from research institutes directed at environmental and sustainability studies, economic policy advice, criminology and law, evaluation studies of development projects, health care studies, to social science. The studies were conducted in the context of the national evaluation protocol SEP for which the institutes have to write self-evaluation reports.[4]

We discuss three institutes, two public and one academic. The two public institutes stated that they wanted to use the SEP protocol (which is not mandatory for them) because they encountered the limitations in other evaluation procedures used before that were not adequate to judge their respective policy oriented missions. By way of contrast, we

1 In this paper we distinguish between academic research organisations that operate in universities and public research organisations that operate outside universities, often focusing on a specific societal sector.

2 SIAMPI was an FP7 project aiming at finding new ways to assess social impact. It stands for Social Impact Assessment Methods for research and funding instruments through the study of Productive Interactions between science and society.

3 Spaapen, J, Van Drooge, L, 2011.

4 Standard Evaluation Protocol (SEP) is since 2003 the national evaluation protocol for all publicly funded research in the Netherlands. It is mandatory for academic research, but most public organisations outside the universities use it too. The current SEP is the third version and runs until 2021.

include also the study of a university institute, which was interested in applying the CRA in view of the increasing attention to broader aspects than academic performance.

While operating in a different context, all three institutes strive to conduct research that is both scientifically excellent and addressing socially relevant issues, but the accents differ. For all three, research topics thus have to be relevant for the scientific community and for society at large. And, the institutes have to find ways to involve stakeholders that are interested in and important for their work. The research agenda of the university research group is primarily geared towards topics in the academic context, while the two public research institutes are primarily directed at topics relevant for the policy makers funding these institutes. It is therefore commonplace that local, regional and national governments are among the stakeholders that want to be involved in developing the research agenda. And, in these networks of stakeholders, power differences pay a role: the one who pays the most – in these two cases the national government – is the one who is likely to have the most influence.

Arguably, the balance between these two goals (producing results for science and society) depends to a certain extent on the policy context in which these institutes operate, including reward systems and local incentives. Most likely, the reward system is different for the public research institutes in this analysis, which main task is to produce reports that are relevant for policy makers. Researchers in the academic institute will be rewarded in the first place for their contributions to the scientific debate, i.e. articles in international journals.

The three institutes also differ with respect to the types of output they produce. The two policy research institutes mainly produce reports addressing issues pertinent to the government or a specific institute acting as customer. But some individual members of the research staff also publish in scientific journals and other academic media very often as a spin off from the research on which the reports are based. By contrast, the academic institute publishes policy oriented reports only as a small fraction of its total output, with a focus mainly on articles in scientific journals, and on books (monographs) and (chapters in) edited volumes, of which a small part of the output is in Dutch.

Our approach aims at gaining more insight about the context by tracing output and getting information about the various stakeholders that are interested in the research produced by these institutes. The method has profited much from the general trend towards open access, which has made it much easier to trace variegated forms of output. As all of these institutes serve public goals, their output follows governmental (and in fact European) policy towards open access, and now as a rule is made publicly available, in print and via websites. The publications of the two policy oriented institutes and of the academic institute may thus reach varied stakeholders both inside and outside academic circles and governments, which is pertinent to the evaluation of the innovation processes in which these institutes take part.

CONTEXTUAL RESPONSE ANALYSIS (CRA)

The method we use is the Contextual Response Analysis. In short, a selected number of publications of each of the three institutes were traced via some generic search engines and specific databases – such as LexisNexis and parliamentary databases, to see who in the environment picks up the results of research as it is published in various ways. The data we collect this way represent the variety of stakeholders of the institute. We classify these data in a radar profile in which the stakeholders are divided in a limited number of categories. The variety of the stakeholder profiles reflects the diversity in interests and topical range of the institute's output.

The CRA also informs about different routes to societal uptake. In the academic institute (social sciences and humanities) routes are often conceived as taking place via popularizations or via reports that address the articulated demands of policy makers, clients and sponsors. However, as our results show, the communication with non-academic users may also be much more indirectly, namely through academic publications. These publications not only serve the academic community but also extensive circles of stakeholders, which numbers are equal or sometimes even much larger than those of stakeholders of reports of the policy oriented public research institutes.

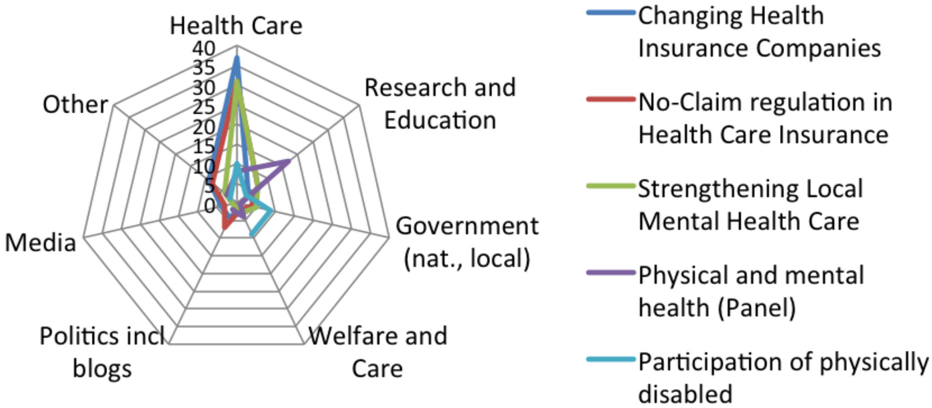


Figure 1 Response Profile for several reports and panels of an institute in Health Care Research: percentage of stakeholders per category (total=100%)

CRA also informs strategic decisions. Contrasts can be drawn, for instance, in the various institutional strategies to achieve and maintain desired levels of societal impact. The awareness of researchers and their institutes for the contextual complexity becomes apparent how the interactions with the environment are managed, some more top down and regular, others bottom up and more ad hoc (Spaapen and Van Drooge, 2011). For policy research institutes we fund more formal meetings and contacts, and examples as different as the maintenance of websites from which reports and their summaries can be downloaded, specific policies for press releases, and monitoring policies for news media attention and social media attention to reports.

In contrast, the academic settings seem to represent a different modality. Firstly, stakeholder connections are loosely organized by the individual researchers or research groups. Also, stakeholder relations are maintained while fulfilling also an academic mission. This does not prohibit stakeholder connections, however. As is also stated in the mission of the institute, there is ample room for researchers to invest in creative and unusual approaches, rather than holding to a demand-driven research agenda. Crucially, it is the policy of this institute to allow for a wide diversity of publication types, including monographs, contributions in newspapers as well as articles in journals. This enables groups and individual researchers to combine various publication channels.

While the method overall gives a good image of the stakeholder context, there are some limitations. The stakeholder profiles that we constructed for the two public research institutes do not fully reflect the organizational characteristics of each of their respective contexts, because some stakeholders pick up research output easier than others. For instance, in some sectors there are many more individually working professionals than larger organizations. Examples are midwifes and general practitioners in health care. It is unlikely that such smaller organizations and individual professionals focusing on the deliverance of health care maintain websites or produce documents referring to reports. In such cases specialized news media (such as professional journals) and specialized knowledge platforms gather, structure and "translate" relevant information and knowledge in order to inform their specific audiences.

Another factor that influences the uptake of publications is the attention in news media given to reports, in particular those with topical issues. Some reports are not only cited more frequently and often also derived more prolonged attention in newspapers, but also seem to attract a more diverse attention from stakeholders because of the received media exposure.

These factors have to be acknowledged in the contextual analysis of these institutes. They also complicate any attempt to a quantitative comparison of the stakeholders in terms of a direct counting of numbers of stakeholders per publication. The profiles are meant to inform assessors with information about the variety of stakeholders in the first place. We realise that in contested fields such as research on global warming or wind energy especially, the diversity of smaller and larger assemblies of professionals or special interest groups, will play a role that is beyond quantitative numbers.

WRAP UP

The participation of variegated stakeholders in innovation processes implies different approaches to evaluation: a wider perspective has to be applied which includes the evaluation of other kinds of output than academic articles and takes into account the interests of all participants in innovative networks.

If we assume that the environment of the identified stakeholders reflect as a whole the goals and missions of the investigated institutes, the information from contextual response analysis can be the basis of a comparison with the mission of the institute. Stakeholders that were expected might be missing, unexpected stakeholders might show up. Also, the question can be addressed about a sufficient balance in the diversity of stakeholders, esp. in contested areas. These results might be used by the board of the institute to modify their policy to manage the relations with stakeholders. Such policy we found in all three institutes, which makes clear that the use of output is not simply a linear trajectory in which use follows production and publication but that the response of stakeholders is part of interactive network processes in which researchers participate.

REFERENCES

The nature, scale, and beneficiaries of research impact. An initial analysis of Research Excellence Framework (REF) 2014 impact case studies, Research Report 2015/01, King's College London and Digital Science, March 2015

Spaapen, J, Van Drooge, L, Introducing 'productive interactions' in social impact assessment, Research Evaluation, 20(3), September 2011, pages 211–218, DOI: 10.3152/095820211X12941371876742

AUTHORS

A.A.M. PRINS
Support in Research Management
Pausgang 4, 9712 EZ Groningen
info@adprins.nl
+31651783381

JACK SPAAPEN
Royal Netherlands Academy of Arts and Sciences
Kloveniersburgwal 29 NL-1011 JV Amsterdam
Jack.Spaapen@knaw.nl

MEASURING THE IMPACT OF A PUBLIC RESEARCH ORGANIZATION ON ENVIRONMENT: A METHODOLOGY BASED ON CASE STUDIES AND AN EXPERT PANEL

LAURENCE COLINET, ARIANE GAUNAND AND PIERRE-BENOÎT JOLY

CONTEXT

Global agriculture will face multiple challenges over the coming decades including confront increased competition for alternative uses of finite land and water resources, adapt to climate change, and contribute to preserving biodiversity and restoring fragile ecosystems (Interagency, 2012). Research and innovation are considered an important solution to address these challenges. In this context, Research Impacts Assessment (RIA) has to take into account a diversity of dimensions, beyond productivity gains, including environmental impacts (Bozeman and Sarewitz, 2011).

The methodology presented in this paper was designed as part of the ASIRPA project at the French National Institute for Agricultural Research (INRA). The ASIRPA project aimed at qualifying and quantifying ex-post the socioeconomic impacts of research at the level of a PRO through the development of a methodological approach based on a series of standardized case studies. 41 cases have been studied to date and selected for their representativeness of INRA's impact pathways. For each case study impacts generated are characterized along five dimensions corresponding to the missions of the PRO and to the international literature (Bornmann, 2013): economic, environmental, political, sanitary, social (Joly et al., 2015). For each of these dimensions, all evidence regarding impacts is collected in the form of local descriptors arising from semi-structured interviews with stakeholders. The intensity of the impact is scored on a scale from 1 (negligible impact) to 5 (major impact) for each dimension. This scoring eases comparability of cases and external institutional communication on impacts. This paper deals with the 1-5 metric for environmental impact.

A set of methodologies aiming at assessing environmental impacts at research program level are available but despite some attempts to combine them (Hermann et al., 2007), there is no consensus on a unified framework. Monetization methods allow for aggregating environmental impacts, but they are costly to implement, and not adapted to a large diversity of environmental impacts. Life cycle analysis, multicriteria analysis, environmental performance indicators methods offer specific advantages and drawbacks, and are adapted to a variety of perimeters. Life cycle analysis (LCA) is a standardized and reproducible method, but encompasses three limitations. First, it is data-, time- and expertise-consuming. Second it generates a single figure that reflects political choices of prioritization of specific environmental objectives. Third, LCA does not account for features of the local context where the innovation is adopted. Fourth, LCA applies to products, not to intangible methods or organizational changes that may arise from agricultural research. Multi-criteria analyses are designed to be decision-making tools and, as far as agriculture is concerned, mostly applied to changes in practices at the farm-level (Galan et al., 2007). They account for the subjectivity of stakeholders' opinions which may change in time. Environmental performance indicators applied ex-post or in-itinere but build only on readily available data (Hermann et al., 2007).

The literature also highlights some barriers related to the indicators of environmental impact. Some difficulties are related to the selection of indicators that can be relevant in a running time and at different scales (local to global biodiversity for instance) (Field to Market, 2012; Walker et al., 2008). Some difficulties are related to the computation of indicators where data are missing or costly (Kelley et al., 2008; Walker et al., 2008). Finally, the heterogeneity of the series of indicators prevents them from being aggregated into a single environmental impact mark.

Given the absence of international consensus on an implementable efficient method to assess environmental impacts of agricultural innovation, there was a need for designing an ad-hoc method.

OUR RESEARCH

The paper presents a methodology, derived from a previous work on political impact (Colinet et al., 2015) to assess the environmental impact of research at the level of a PRO. Our methodology consists in building a generic metric based on the results of standardized case studies submitted to a panel of experts (Cohen et al., 2015; Ruegg and Feller, 2003). That judging metric should be equally credible to the metrics of the other 4 dimensions (economic, political, sanitary, social), it should be relevant for all the case studies released on INRA's impact by the ASIRPA team, and beyond that, comprehensive and general enough so as to allow the regular addition of a diversity of cases.

A first step of desk research on the grey and academic literature (EMBRAPA, 2014; Field to Market, 2012; CGIAR: Walker et al., 2008; Hazell and Haddad, 2001; Maredia and Pingali, 2001; Renkow and Byerlee, 2010; Tekes: Luoma et al., 2011Global Environmental Facility : van den Berg and Todd, 2011) reveals that agricultural research organizations often disentangle environmental impact into localized effects, global effects, and pressure on resources. We built on these experiences to draw an initial framework made of four categories:

- Local or national impact related to pollutions and destruction of ecological compartments (1)
- Global impact related to issues with international commitments: biodiversity (2) and climate change (3).
- Impact on resources consumption (4)
- We first considered the 26 case studies (out of the 34 cases available at the start of that work) where environmental impact was considered to be significant. For each case, all the descriptors expressed by the stakeholders were affected to one of the four categories. This information was reported in four tables.

We gathered a panel of five experts in the environmental policies design and implementation from public institutions: two different ministries (in charge of environment and agriculture), one extension service (Ademe), and one research institute (MNHN). All experts were French, and had a research background. We wanted to keep the panel small to promote interactions and consensus, while trying to somewhat overlap competences. All experts were knowledgeable about the agricultural and environmental policies over the past decades.

The consultation was carried out in two steps. Experts were asked to first remotely and individually review the evidence collected in the cases and rate the impact of each case. The rating was to be made on a 1 to 5 scale on each of the four category; negative impacts are to be mentioned. The expected judgment consisted on comparing, for each case on each subdimension, the achieved intensity of impact as compared to the maximum possible impact. The impact judgment was to be made independently from the contribution of INRA to that impact since this last feature was already accounted for. Then we organized two meetings where experts could confront their views on their rating and judgment criteria. The objective of the first meeting, held in February 2015, was not to build a consensus to reach one single mark for each case, but to elicit the criteria on which each expert based his/her judgment, and what was to them the relative value of these criteria. This first meeting enabled us to revise the analytical framework and to inductively derive relations between descriptors of impact of each sub-category and marks on a 1 to 5 scale. The second meeting of the panel, held in September 2015, aimed at validating and consolidating that grid, as well as discussing a procedure for aggregating the 4 sub-dimensions into a single mark of environmental impact.

OUR RESULTS

THIS ORIGINAL PROCEDURE YIELDED THREE TYPES OF RESULTS.

Learning on specific features of environmental impacts that influence its judgment has been reached. First, and contrary to political impact (Colinet et al., 2015), there may be interferences between subdimensions of environmental impact: an innovation can have a positive impact on climate change while affecting negatively resource consumptions. This remark lead to cautiously design the aggregation procedure proposed below. Second, the ASIRPA approach focuses on assessing ex-post achieved impact, instead of potential impacts; still, given the irreversibility of environmental impacts, uncertainty, complexity and systemic effects, including on long time scales (future generations) have to be accounted for. This leads to consider both the quality of the research outputs, the socio-technical dynamics affected, and the cyclical nature of the environmental threat. This remark leads to include criteria related to sustainability, permanence of context and performance of the research outputs. Other learning arose from the expert panel, notably regarding the curative versus systemic effects of the research outputs, and spurred to refine our definition of environmental impact and fine-tune the choices and limitations of the scaling grid released.

A scaling grid has been designed that allows to objectively self-assess (based on stakeholders' opinions collected) on a 1 to 5 scale the environmental impacts of all types of research outputs from an agricultural public research organization. Considering the diversity of case studied, we claim that the method and its grid are relevant for any agricultural PRO. It comprises four subdimensions of environmental impact, in addition to a transversal grid. The transversal grid rates the originality and quality of the research outputs, the scale of adoption and the systemic nature of the impacts observed. The four subdimensions are the one initially considered: pollutions and destruction of compartments, biodiversity, climate change, resources consumption. For each of these subdimensions, four criteria are considered: the importance and gravity of the stakes/problems; the originality and quality of the research outputs as far as the biodiversity or climate change or resources or pollutions are concerned; the geographical scale of adoption, as compared to the potential perimeter which is relevant for the considered subdimension; the specific impacts on biodiversity or climate change or resources or pollutions. This grading scale considers impact in terms of the environmental performance of the research outputs released and the intensity of their adoption. The procedure enables liberating from the subjectivity of expert judgments on a sample of cases by releasing an objective and standing-alone grading scale that can be implemented in a self-assessment process. This result can produce an overview of the environmental impacts of a portfolio of innovations at the level of an organization, and can be adapted for other types of impacts. This metric can be used to teach lessons at the level of a PRO through the regular addition of case studies.

Two proposals have emerged for aggregating the transversal grid and the four subdimensions of environmental impact into a single score, which fulfills requirements for assessment at the level of INRA as a whole. The first one relies on an algorithm to compute a weighted average of the subdimensions marks. It was designed in order that the final ranking of cases on the grid would be discriminant, would deliver a dynamic incentive message for INRA, and would neither penalize "specialised cases" that intensely impacts a single subdimension that holds great stakes (e.g.: biodiversity), nor "polyvalent cases" that affect in a more moderate ways all the subdimensions of environmental impact. Another poorly exploited approach would consist in invoking an integrative concept. This concept could be related either to the Sustainable Development Goals, or to the ecosystemic services defined by the Millenium Assessment Goals. A matching of each subdimension's score with that integrative concept would provide us with a single mark of environmental impact.

Further steps involve rescoring all the ASIRPA cases available to date in that grid in order to test its robustness. It is foreseen that the expert panel will be regularly consulted in order to adapt the scaling grid according to new types of environmental impacts discovered along cases or new missions assigned to the organization.

REFERENCES

Bornmann, L., 2013. What is societal impact of research and how can it be assessed? a literature survey. Journal of the American Society for Information Science and Technology 64, 217–233.

Bozeman, B., Sarewitz, D., 2011. Public value mapping and science policy evaluation. Minerva 49, 1–23.

Cohen, G., Schroeder, J., Newson, R., King, L., Rychetnik, L., Milat, A., Bauman, A., Redman, S., Chapman, S., 2015. Does health intervention research have real world policy and practice impacts: testing a new impact assessment tool. Health Research Policy and Systems 13, 3. doi:10.1186/1478-4505-13-3

Colinet, L., Gaunand, A., Joly, P.-B., 2015. Measuring the impact of a public research organization on public policies: a methodology based on case studies and an expert panel. Presented at the ImpAR Conference 2015, Paris, France.

EMBRAPA, 2014. Social report 2013. Embrapa, Communication Secretariat, Strategic, Management Secretariat, Brasilia.

Field to Market, 2012. Environmental and Socioeconomic Indicators for Measuring Outcomes of On-Farm Agricultural Production in the United States.

Galan, M.B., Peschard, D., Boizard, H., 2007. ISO 14 001 at the farm level: Analysis of five methods for evaluating the environmental impact of agricultural practices. Journal of Environmental Management 82, 341–352.

Hazell, P., Haddad, L., 2001. Agricultural research and poverty reduction. Food Agriculture and the Environment, International Food Policy Research Institute. Technical Advisory Commitee of the Consultative Group on International Research in Agriculture Discssion paper 34.

Hermann, B., G., Kroeze, C., Jawjit, W., 2007. Assessing environmental performance by combining life cycle assessment, multi-criteria analysis and environmental performance indicators. Journal of Cleaner Production 15, 1787–1796. doi:10.1016/j.jclepro.2006.04.004

Interagency, 2012. Sustainable Agricultural Productivity Growth and Bridging the Gap for Small-Family Farms: Interagency Report to the Mexican G20 Presidency. Interagency: Bioversity, CGIAR Consortium, FAO, IFAD, IFPRI, IICA, OECD, UNCTAD, Coordination team of UN High Level Task Force on the Food Security Crisis, WFP, World Bank, and WTO.

Joly, P.-B., Gaunand, A., Colinet, L., Larédo, P., Lemarié, S., Matt, M., 2015. ASIRPA: A comprehensive theory-based approach to assessing the societal impacts of a research organization. Research Evaluation 24, 440–453. doi:10.1093/reseval/rvv015

Kelley, T., Ryan, J., Gregersen, H., 2008. Enhancing ex-post impact assessment of agricultural ersearch: the CGIAR experience. Research Evaluation 17, 201.

Luoma, P., Raivio, T., Tommila, P., Lunabba, J., Halme, K., Viljamaa, K., Lahtinen, H., 2011. Better results, more value: a framework for analysing the societal impact of research and innovation. Tekes, Helsinki.

Maredia, M., Pingali, P., 2001. Environmental Impacts of Productivity-Enhancing Crop Research: A Critical Review. TAC Secretariat Food and Agriculture Organisation of the United Nations. Consultative Group on International Agricultural Research Technical Advisory Commitee, Durban, South-Africa.

Renkow, M., Byerlee, D., 2010. The impacts of CGIAR research: A review of recent evidence. Food policy 35, 391–402.

Ruegg, R., Feller, I., 2003. A Toolkit for Evaluating Public R&D Investment: Models, Methods, and Findings from ATP's First Decade. (Grant/Contract Report). National Institute of Standards and Technology, Gaithersburg.

van den Berg, R.D., Todd, D., 2011. The full road to impact: the experience of the Global Environment Facility Fourth Overall Performance Study. Journal of Development Effectiveness 3, 389–413. doi:10.1080/19439342.2011.562302

Walker, T., Maredia, M., Kelley, T., Rovere, R.L., Templeton, D., Thiele, G., Douthwaite, B., 2008. Strategic Guidance for Ex Post Impact Assessment of Agricultural Research. Report prepared for the Standing Panel on Impact Assessment, CGIAR Science Council, Science Council Secretariat: Rome, Italy.

AUTHORS

LAURENCE COLINET
INRA, 147 rue de l'Université
75338 Paris Cedex 07, France

ARIANE GAUNAND
INRA, Délégation à l'Evaluation
147 rue de l'Université, 75338 Paris Cedex 07, France
and
Université Paris-Est
LISIS 77454 Marne-la-Vallée, France
tel: 0033 (0)1 42 75 91 06
ariane.gaunand@paris.inra.fr

PIERRE-BENOÎT JOLY
Université Paris-Est
LISIS 77454 Marne-la-Vallée, France

EVALUATING THE IMPACTS OF NEW ZEALAND'S CROWN RESEARCH INSTITUTES – FRAMEWORKS, FORUMS AND FOSTERING DEVELOPMENTAL EVALUATION IN RESEARCH PROGRAMS

T. A. WILLIAMS, T. WHITE, W. KAYE-BLAKE, G. GREER AND H. PERCY

INTRODUCTION

Recipients of public and private investments in RS&T face increasing expectations that their research endeavours will deliver innovations that will, in turn, have an impact on economies, society, and the environment. For their part, funders of research addressing complex problems that challenge New Zealand seek reassurance that their investments are generating an adequate return, in terms of both impact and science excellence. Armed with this evidence, investors seek to make better informed investment decisions in order to protect and enhance the lives of New Zealanders and ensure, as a nation, we are well positioned to make an effective contribution to global challenges and opportunities.

While evaluation of investments in education, health and international aid has been practised since mid last-century, Monitoring & Evaluation (M&E) methods have only recently been applied to RS&T. Early experiences confirm that similar challenges experienced by M&E practitioners in these portfolios consistently arise when evaluating RS&T, including issues around time lags, attribution and evaluative capacity.

In New Zealand, 7 Crown Research Institutes (CRIs) are responsible for promoting and facilitating the application of the results of research and technological developments, and having regard to the principles of the Treaty of Waitangi (Crown Research Institutes Act 1992)[1]. This paper describes the evolving policy setting for CRIs as well as M&E expectations of these diverse organisations. It describes the response to these expectations in 2 CRIs charged with delivering innovation to the agricultural, horticultural and seafood sectors Ð The New Zealand Institute for Plant & Food Research Limited (PFR, www.plantandfood.co.nz) and AgResearch Limited (www.agresearch.co.nz). It also describes a new network, iPEN (the Impact Planning & Evaluation Network), that has been established across the CRIs to share learning and resources as well as challenges experienced in integrating M&E theory and practice into research programs. The paper provides evidence of 2 new approaches to evaluation in PFR and Agresearch that seek to embed developmental evaluation in research programs while supporting a co-innovation approach to delivering impact.

THE NEW ZEALAND POLICY SETTING FOR CRIS

The CRIs are science research businesses owned by the Crown (the New Zealand Government). Collectively, they are the largest dedicated providers of science research in New Zealand. More than 3600 people work in the 7 CRIs, which are organised around providing solutions to New Zealand's critical issues for the economy, environment and society. The CRIs undertake blue-sky and applied science and technology research and development. Their clients include central and local government and private sector markets in New Zealand and abroad (https://careers.sciencenewzealand.org/crown-research-institutes).

Each CRI has a Statement of Core Purpose that outlines the organisation's purpose, outcomes, scope of operations (including the key sectors on which it should focus its activities) and operating principles. Essentially, CRIs must remain financially viable, develop strong, long-term partnerships with key stakeholders and work with them to set research priorities that are well linked to the needs and potential of their end users (http://www.plantandfood.co.nz/file/pfr-scp.pdf).

ACCOUNTABILITY FRAMEWORK

Over the nearly 25 years that CRIs have been in operation, these organisations have submitted annually a Statement of Corporate Intent, a confidential Business Plan and a publicly available Annual Report to their two Shareholding Minister Ð the Ministers of Science and Innovation, and Finance. A series of key performance measures has also been

1 The Treaty of Waitangi is New Zealand's founding document that outlines a broad statement of principles on which the British and MÐori (indigenous people of New Zealand) made a political compact to build a [single] government in New Zealand.

established for annual reporting, including impact case studies, metrics and stakeholder surveys of various aspects of performance.

In 2011 Core Funding was introduced to increase stability and efficiency in the New Zealand Research, Science & Innovation system (RS&IS). As well, a set of generic performance indicators across all CRIs was introduced, including a series of 4-year rolling reviews for each CRI on diverse aspects of their operations. These have provided sector-level insights into the contribution the CRIs make to New Zealand's economy, society and environment.

UNDERSTANDING POTENTIAL AND DELIVERED IMPACTS OF SCIENCE

In the last 6 months the New Zealand Government has released its National Statement of Science Investment (NSSI) for 2015/25. In the foreword, the Minister of Science and Innovation states: "The Government believes excellent, high impact science is fundamental to our ability to achieve excellent economic, environmental, social and cultural outcomes for New Zealand" (p.4, http://www.mbie.govt.nz/info-services/science-innovation/pdf-library/NSSI%20Final%20Document%202015.pdf).

The document takes a systems view and outlines a vision by 2025 of "A highly dynamic science system that enriches New Zealand, making a more visible, measurable contribution to our productivity and wellbeing through excellent science" (p.10, italics added). As well it outlines the precise role of Government, CRIs and other public investments in science in a "horizons-based model for thinking about public science investments" (p.31, Fig. 1). CRI Core Funding is located towards the applied end of the horizon where significant leverage can be gained from proven ideas, suggesting its primary focus needs to be in delivering impact, although a recent rebranding of this fund to the `Strategic Science Investment Fund' has refocused attention on underpinning platforms of research to maintain capability for New Zealand. New funding instruments, the National Science Challenges, are located closer to the start of the pipeline, where new pan-sector are to be addressed while the Ministry of Business, Innovation and Employment (MBIE, a significant broker of contestable funds for the CRIs) is positioned mid-way in the pipeline with a focus on higher risk science with longer term impact.

This specification of the roles of investments and the Government's continued encouragement of increased levels of private sector investment in R&D clearly signals an increased expectation of impact delivery for CRIs, but not at the expense of science excellence.

To guide analysis of the performance of the RS&IS the New Zealand Government has undertaken a Domain Plan to identify data needs as well as current data sources and gaps. A Science Intel Database is under construction to help address questions about the performance of the RS&IS. An annual system performance report is also being developed to provide a "point-in-time snapshot of the performance of science and innovation in New Zealand". It will cover measures such as R&D intensities, research quality and commercialisation outcomes, public investment in science and innovation, institutional performance, business innovation measures, and public engagement with RS&T (p.55).

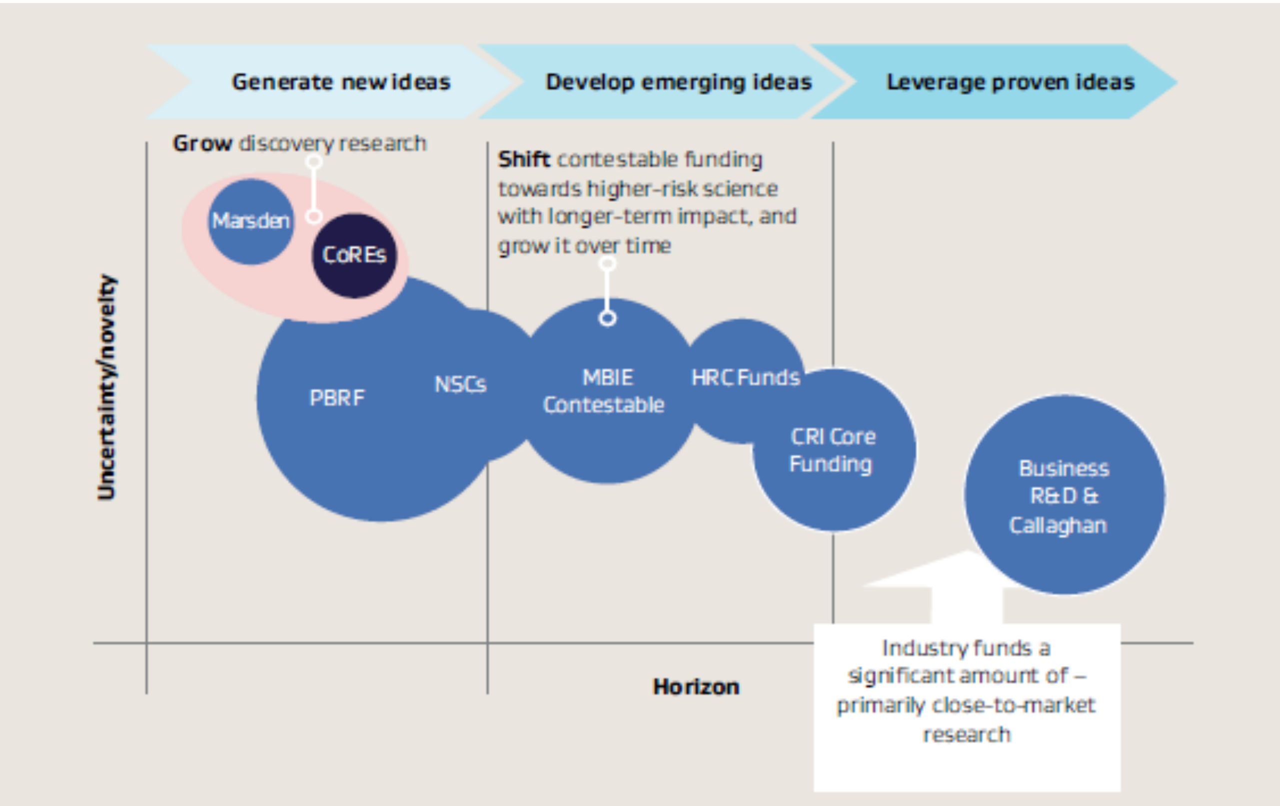


Figure 1. A horizons-based model for thinking about public science investment (MBIE 2015, p.31).

CAPTURING THE FULL RANGE OF VALUES CREATED BY RS&T PROGRAMS

As mentioned earlier, frameworks for evaluating the impacts of public investments in research are emerging. Jaffe (2015) reviews a number of them, identifies a set of metrics and indicators that covers major dimensions of public research impacts (Table 1), and notes cross-cutting issues when attempting to evaluate long-term impacts of research.

Table 1. Dimensions of public research impact (adapted from Jaffe 2015).

Economic:	1. New or improved products or services
	2. Reduced operating cost or reduced commercial risk
	3. Increased wages or improved job opportunities
Environmental:	4. Reduced pollution or other anthropogenic environmental impact
Public policy:	5. Improvement of public policy or of the delivery of public services
Capability:	6. Enhancement of the scientific and technological capabilities of the work force
Social:	7. Improved morbidity and mortality, or reduction in the cost of maintaining health
	8. Increased knowledge and interest in science
	9. Reduction in real or perceived communal risk
	10. Enhancement of NZ international reputation, or contribution to sustainable development in other countries
	11. Enhancement of social, cultural or community values

MBIE recently established a reduced set of generic indicators for application across the CRIs. These indicators focus on 5 high-level dimensions of performance relevant to all CRIs that have been normalised to generate comparative data (Fig. 2).

Figure 2. Indicators used by the Ministry of Business, Innovation and Employment to evaluate CRI performance.

End-user collaboration	Revenue per FTE from commercial sources. **Quarterly.**
Research collaboration	Publications with collaborators. **Quarterly.**
Technology and knowledge transfer	Commercial reports per scientist FTE. **Quarterly.**
Science quality	Impact of scientific publications. **Annually.**
Financial indicator	Revenue per FTE. **Quarterly.**

As well, CRIs are required to publish at least 9 case studies in their Annual Reports, describing innovations that align with their Statements of Core Purpose as well as the quantifiable impact on industry or sectors from these innovations. Over the last 2 years MBIE has encouraged CRIs to develop their own performance frameworks and indicators that reflect key strategic initiatives and to document these in the CRIs' publicly available Statements of Corporate Intent. This has seen CRIs increase their focus on M&E and particularly mechanisms to ensure they are `investing for impact' efficiently while maintaining science capacity and excellence.

`Investing for impact' is, in turn, leading PFR and AgResearch to explore the practical benefits of adopting co-innovation approaches to accelerate the pace of innovation and delivery of impact to enable more profitable growing and farming practices. The process is being guided by nine principles of co-innovation that have been adapted and applied by a research program, 'Primary Innovation'[2], in innovation projects traversing key industries in New Zealand's primary sector. Results are demonstrating that developmental evaluation methods are best suited to co-innovation. These methods provide sufficient adaptability and flexibility to capture the increased interactions and collaborations that the MBIE indicators in Fig. 2 seek to measure. As a consequence, they help to meet accountability expectations of stakeholders.

Responsibility for evaluating the non financial performance of CRIs falls to a range of specialists within the CRIs, from science strategy managers, to communications specialists to dedicated Impact Evaluation Managers. In the last 18 months these specialists have formed a new network, iPEN (the Impact Planning and Evaluation Network). The purpose of the network is to share knowledge, resource and experiences amongst the CRIs and with key funding ministries, including MBIE, the Ministry of Primary Industries, and the Ministry for the Environment.

EMERGING APPROACHES AT PLANT & FOOD RESEARCH AND AGRESEARCH AND THE CASE FOR DEVELOPMENTAL EVALUATION

The increased focus on the impact contributed by CRIs and other funders-providers in the New Zealand RS&IS has had a number of effects on the way RS&T is evaluated in the CRIs. These organisations have refined their quantitative and qualitative performance measures and have linked them more closely to their strategy in order to better tell their performance story. For example, PFR has created a performance framework that identifies areas of strategic focus across the entire organisation as well as a set of 40+ indicators (including the MBIE generic indicators). A Growing Futures website (www.growingfutures.co.nz) elaborates on PFR's impacts and science excellence. Novel approaches to putting a value on the total outputs and impacts of CRIs are also emerging. PFR has undertaken workshops with five of its key sectors in which industry and research representatives discuss the drivers of innovation and the contribution of PFR's research to industry performance over the last 10-15 years. This innovative approach to evaluation has provided a broader context within which to locate and better understand the meaning of program-level evaluations conducted using methods such as Cost Benefit Analyses. The discursive workshop approach also revealed the importance of trust and strong personal relationships between industry and science representatives in order to support adaptability, flexibility and reflection.

A greater focus on evaluation, particularly developmental evaluation, to inform learning and further enhance impact through co-innovation

2 The Primary Innovation is an MBIE-funded contract (CONT-30071-BITR-AGR) that is encouraging and sharing ideas to foster co-learning and co-innovation to bring greater economic benefit and a more sustainable future for New Zealand (www.beyondresults.co.nz/PrimaryInnovation/Pages/default.aspx).

is also emerging and shaping the way RS&T is undertaken. In 2013 AgResearch launched an Adoption and Practice Change Roadmap (http://www.beyondresults.co.nz/About/Pages/default.aspx) and a series of initiatives to enhance the impact of agricultural R&D for the pastoral, agri-food and agri-technology value chains. Activities include case studies, the development of planning tools, building capability in M&E through a network of champions, identifying key skills and competencies, including the development of the innovation brokering function, and supporting the role of advisory groups to bring about practice change.

At a program level, this initiative has identified the need to encourage stakeholders (including researchers) to work together to develop impact pathways, engage more actively in discussions about research priorities, identify data they may contribute and involve stakeholders in interpreting research findings. The need for increased skills within research programs in evaluation, facilitation, data visualisation, innovation brokering and reflection has also been recognised. Indicators jointly identified and applied by researchers and stakeholders to capture the value generated by co-innovation are still under development. This focus on developmental evaluation has the joint purpose of analysis and learning as opposed to accountability, advocacy and resource allocation.

At a systems level, the Primary Innovation program is helping to locate systemic barriers to innovation and opportunities for structural reform at the institutional level. In particular, a need has been identified for funders of RS&T to be more flexible in their expectations of project management towards contracted outputs and rigid milestone delivery and rather accommodate a more dynamic and flexible pathway to impact.

CONCLUSION

The extent to which developmental evaluation can increase the capacity of researchers and stakeholders to co-innovate and, in turn, enhance the delivery of impact from RS&T is yet to be determined. Two initiatives in the CRIs to increase engagement between stakeholders have demonstrated insights from reflecting on the contribution of RS&T to industry-level innovation in a workshop setting, and embedding evaluation capacity within research programs to enhance learning. At a systems level, opportunities exist for funders to develop more flexible reporting frameworks, and require program logics and M&E plans in research proposals to incentivise these behaviours within research organisations. The identification of indicators that better capture diverse sources of value and the data sets required to measure those indicators is a work-in-progress at both system, organisational and program levels.

The current frameworks used to evaluate the performance of the CRIs and the impacts they deliver have suited the first 25 years of their operation. Developmental evaluation appears well placed to generate evidence of the diverse interactions, networking, trust-building and learning created by co-innovation in practice in the future. Over time, as the theory and practice of co-innovation evolve in New Zealand, new types of indicators, data and systems will emerge that better reflect the value generated by RS&T programs.

REFERENCE

Jaffe, A. B. (2015): "A framework for evaluating the beneficial impacts of publicly funded research". MOTU Note No. 15. Wellington, New Zealand (http://motu.nz/)

AUTHORS

T. A. WILLIAMS
Contact Author
The New Zealand Institute for Plant & Food Research Limited
Private Bag 4704, Christchurch, New Zealand.
Ph +64 3 940 3949

T. WHITE
AgResearch Limited
Ruakura, Hamilton New Zealand
Ph +64 7 838 5763

W. KAYE-BLAKE
PricewaterhouseCoopers
Mail Centre 8140, Wellington, New Zealand
Ph +64 4 462 7033

G. GREER
Agribusiness Economics Research Unit
P O Box 85084, Lincoln University, Lincoln 7647, Christchurch, New Zealand. Ph +64 3 3 423 0376

H. PERCY
AgResearch Limited
Ruakura, Hamilton New Zealand
Ph +64 7 838 5107

EVALUATION OF RESEARCH INSTITUTIONS OF THE NATIONAL ACADEMY OF SCIENCES OF UKRAINE: OLD AND NEW APPROACHES.

IGOR YEGOROV

INTRODUCTION

R The National Academy of Sciences of Ukraine (NASU) is leading research organization of the country. It includes 168 research organizations, which form 3 sections and 14 departments according the distribution of institutes to scientific disciplines. Academy has 21.3 thous. researchers, it's total budget was 2683 million Hryvna Ukr. in 2014 (less than 110 million Euros according to the market exchange rate). Bulk of the money comes from the state budget. Almost all Ukrainian journals from Thomson-Reuters data base are published by the NASU.

NASU has high reputation in the country. However, but it has preserved some features from the Soviet period, which provoke criticism in society and from abroad. Most critics refer to the obsolete managerial system and insufficient transparency in decision-making processes, including distribution of research funds.

CURRENT PROCEDURE OF EVALUATION

Proper evaluation of research potential of NASU institutes has to be a key element (and precondition of reforms). Current procedures of evaluation have been formalized in 1998 by the special Decision N 469 of the Cabinet of Ministers of Ukraine 'On approval of the state certification of scientific research (scientific and technical) institutions'. Several changes have been made since this time but almost all of them were not radical.

Evaluation is conducted to assess the effectiveness of institutions and to show how results of their activities correspond with the state priority areas in science, technology and innovation. As a result of evaluation, institute has to receive a document about the state certification. It is important to stress that all R&D institutions, which receive state support, are subjects of certification. As a result of certification, institute could be included into the State Register of scientific institutions. This, in turn, opens the way for applying to the money from the state funds and to obtaining some kind of tax privileges. Evaluation takes place once in 5 years. Presidium of the NASU (the highest governing body of the Academy) is responsible for the evaluation. Evaluation itself includes: survey of scientific organizations and supporting technical institutions by the special commission from the specialists from inside and outside of the institute; evaluation at the level of departments with possibility to use some extra information and checking of surveyed forms by the administration of the institute; multidisciplinary expertise (at the level of Presidium) and ranking of research institutions.

Survey includes information on scientific, technical, and teaching staff (number of employees, who perform R&D, the number of doctors and candidates of sciences, graduate and doctoral students); description of the main results and general scope of scientific activities (number of theses, publications, books, encyclopedias and dictionaries, textbooks (manuals), articles in scientific journals, including journals, participation in the international scientometric databases and so on); assessment of practical value of scientific and technical activities for specific sectors and the national economy as a whole, completed applied research projects on which conducted experimental development and some other information. Some indicators are measured quantitatively. These indicators include level of financing of design and technological projects, aimed at creation of prototypes (thous. UAH); volume of scientific and technical services (thous. UAH); national and international recognition of research results (number of received awards, including state and international awards and grants of the President of Ukraine, the Cabinet of Ministers of Ukraine, the National Academies of Science); number of foreign grants; membership in professional scientific societies and, foreign academies. Special procedures of generalization of individual indicators were developed to receive one-figure-estimate, which could be used for ranking procedure. However, in reality, this procedure has not been used in a strict way. No institutes have been closed on the base of evaluation.

NEW APPROACH TO THE EVALUATION OF RESEARCH ORGANIZATIONS

In 2015, it was a decision to change the procedure of evaluation of the institutes of the National Academy of Sciences in the context of general reform of Ukrainian scientific system. New evaluation has to be based on utilization of international experience and national and international indicators. Procedures of evaluation have to be transparent and democratic. Exclusion of conflict of interests has to be an important feature of this procedure. On the other hand, there have to be possibility to appeal results of the evaluation from the side of research organization. New procedure has to be more flexible one (no single indicator for ranking). As to the experts, involvement of external evaluators is a key precondition of success. Ukraine has decided to utilize German experience of Lebnitz Association as of the similar organization to the National Academy. Evaluation team consists of 3 groups: expert group; permanent expert committee on a relevant field of science; permanent evaluation committee of the National Academy of Sciences of Ukraine. At the first stage, the expert group (first-level review board) evaluates the scientific activities of the institution. The members of the group inspect the institution's activities, analyze the inquiry form filled by the institution beforehand, verify whether the materials submitted by the institution are unbiased, and prepare their conclusion according to the selected criteria. At the second stage, the Permanent Expert Committee on a Relevant Field of Science (second-level review board) prepares a presentation on the institution activities in accordance with the report of the first-level group and after consultations with the institution. The second-level review board conveys the conclusion of the first-level group to the institution. The institution can make a statement concerning this conclusion. At the third stage, the Permanent Evaluation Committee of the National Academy of Sciences of Ukraine (third-level review board) considers the presentation of the second-level board, the conclusion of the first-level group, and the statement of the institution. The third stage of the evaluation should result in the report of the third-stage review board that should evaluate the scientific activities of the institution and contain recommendation on its further financing. The report of the third-stage review board should be based on the results of the first-level and second-level evaluation stages. The institution has the following opportunities to take part in the evaluation procedure: prior to the selection of experts of the first-level review board by the second-level review board; the institution can propose a list of main research fields to be covered by the evaluation procedure; the institution can propose experts in these research fields according to the criteria that determine a potential conflict of interest; following the selection of experts of the first-level review board by the second-level review board, the institution can comment on whether the experts cover the research fields named by the institution; the institution can comment on whether it sees a potential conflict of interest among the experts selected. In case the second-level review board and the institution fail to reach an agreement after the discussion of the comments, the final decision should be made by the first-order review board. The institution obtains a mandatory copy of the first-level review board conclusion from the second-level review board and it is obliged to prepare its statement concerning the conclusion of the first-level review board.

Criteria for evaluation of the quality of work and potential of an institution by the first-level review board: development of the institution in previous years and its research strategy for the next years; scientific results; scientific events and public outreach; appropriateness of facilities/financial provision. Special attention is paid to the collaboration and networking (several positions are usually considered).

Key quantitative indicators of evaluation are the following: number of publications (depending on the publication culture of the subject area, in particular in peer-reviewed journals, at peer-reviewed conferences, in monographs etc.); number of commercial property rights and patents, the number of consulting contracts and expert reviews; the amount of third party funds raised for research, consulting, services, etc.; the income from commercial activity, lease. Other quantitative indicators could be also included into evaluation procedure. Quality assurance is provided by the a) internal quality management at the institution and b) by assessment of the institution by the relevant Department of the NAS of Ukraine.

Strategic significance of the institution is determined by the answering the following questions as a result of evaluation:

Is the institution of strategic significance: for the further development of a certain special field and its environment? as a hub for specialists or regional clusters? for the further development of fields of technology, information and other services, consulting, social-political tasks? for the profiling of programs of the NAS of Ukraine?

During the first stage of the evaluation Ukraine will need an assistance in provision of independent experts for evaluation and participation in evaluation procedures and organization of consultations on introduction of assessment procedures and creation of the pool of experts.

New procedure was approved at the very beginning of the 2016, and the evaluation of the first 10-15 institutes will be made during the summer – beginning the autumn of 2016. First results will be presented at the Conference.

AUTHOR

IGOR YEGOROV
Institute of Economy and Forecasting
National Academy of Sciences of Ukraine
26 Panasa Myrnogo st.,
Kiev, 01011, Ukraine
(380-44) -2801402

USING AN ASSESSMENT OF 'COMPLICATED' AND 'COMPLEX' CHARACTERISTICS TO DETERMINE EVALUATION DESIGN OF INNOVATION POLICIES

JONATHAN COOK

CONTEXT AND OBJECTIVES

Policies in the field of Research, Technology and Innovation (RTI) are often not associated with linear processes of cause and effect. This partly reflects the nature of innovative activities, whereby the results and the ways in which firms apply these can be unclear at the outset, in particular where innovation takes place at the frontiers of knowledge (Hof et al., 2012). In addition, the increasingly 'open' and collaborative way in which innovation is undertaken can mean that some of the benefits of RTI policies are indirect and unintended, as results are diffused through the innovation network system (Jordan, 2010), e.g. through knowledge spillovers. Moreover, the benefits of RTI policies can be unevenly distributed, with small numbers of beneficiaries/actors reaping the vast majority of the rewards (Cook et al., 2013).

The specific ways in which some RTI policies are designed create challenges for evaluators seeking to assess cause and effect. For example, reflecting the iterative and collaborative process of innovation, policies can involve multiple components or partners. Other policies may involve support that is highly tailored to specific contexts and circumstances such that no 'standard' intervention exists. The rise of 'demand-side' policies has also resulted in the need to consider a broad set of inter-relationships between different institutions within a system. Alongside these challenges, policy-makers' expectations of evaluation are high. In particular, there is a desire for evaluation to place a 'value' on policies and programmes to inform future decisions and investments, ideally through the use of experimental or quasi-experimental approaches. How can evaluators best respond to these challenges and expectations?

Evaluation literature has suggested that the characteristics of interventions can be used to inform evaluation design, with Rogers (2008) drawing a distinction between aspects of interventions that can be categorised as 'simple', 'complicated' and 'complex'. Rogers (2008) also illustrated how programme theory can be used in complicated or complex situations without resort to "messy" logic models. For interventions that exhibit features that are complicated or complex (or where the population of beneficiaries is small), counterfactual-based approaches to evaluation may be inappropriate. In these cases, theory-based evaluation approaches may be used (Rogers, 2007; Weiss, 2000; White, 2009). White and Phillips (2012) described a range of these theory-based techniques, including contribution analysis, process tracing and realist synthesis, which can be used to assess the extent to which interventions have brought about outcomes.

In this paper, we draw on our recent evaluation studies relating to RTI policies to examine the extent to which different policy interventions exhibit the 'simple', 'complicated' or 'complex' characteristics set out by Rogers (2008). We then describe how these characteristics can be used as determinants of appropriate evaluation design, and the role of programme theory as a tool to inform evaluation. Finally, we examine the extent to which evaluation designs for RTI policies with 'complicated' and 'complex' characteristics are likely to meet policy-makers' expectations of valuing the contribution of RTI policies to the economy.

The paper draws on the experiences of several recent evaluation studies that we have undertaken covering evaluation scoping studies and programme evaluation assignments. They are focussed on business innovation in the UK and the EU, and the mix of interventions includes: single company R&D grants; collaborative R&D grants; investments in new RTI infrastructures that seek to bridge the gap between research and businesses; and demand-side innovation policies.

ASSESSING THE CHARACTERISTICS AND PROGRAMME THEORIES OF POLICIES

Drawing on the classification of issues identified by Rogers (2008), Table 1 sets these into the context of RTI policies, and extends the range of issues to cover other aspects that we have found to be important. In summary, these aspects are as follows (drawing on Rogers, 2008):

- The nature of implementation and engagement takes account of the extent to which multiple partners are involved in delivery or as part of innovation partnerships.
- Simultaneous causal strands mean that two or more routes to outcomes are required to occur for an intervention to work, such as technical success of an R&D project along with the develop-

ment of innovation capacities to take the output to market.

- Alternative causal strands are subtly different as they mean that there could be more than one causal route for a programme/ policy, which can be particularly relevant where the intervention is highly tailored.
- Timescales to outcomes can be long for RTI policies, and a further issue is the extent to which they may vary across a policy, e.g. with recipients of an intervention achieving outcomes over different timeframes.
- Policy objectives may be focussed on economic issues (e.g. related in some way to growth), though may also cut across a range of issues (e.g. economic, societal and system).
- Non-linearity of outcomes reflects that an initial effect may result in a feedback loop that brings about further rounds of effects, i.e. acts as a tipping point.
- Outcomes can be pre-identified based on known or at least anticipated relationships (i.e. non-emergent outcomes) or can be dependent on interactions between different organisations, and sometimes be unpredictable (i.e. emergent outcomes).

This classification provides a set of determinants for evaluation approaches. For example, the proliferation of characteristics that are 'simple' will lend weight to experimental or quasi-experimental approaches, and where 'complex' characteristics are significant, theory-based (or alternative) approaches will be required. For policies with 'complicated' characteristics, there may be a choice or a mix of experimental/quasi-experimental and theory-based approaches, depending on other parameters and key evaluation questions.

The final choices of evaluation design will then require further consideration and refinement. For example, an R&D grants policy may have mainly simple characteristics (e.g. on implementation and engagement with the policy, causal strands, and policy objectives), and a quasi-experimental approach can be adopted. However, the high degree of skewness in outcomes, the potential variability in timescales to outcomes, and the alternative policies available in the wider RTI landscape can pose challenges to analysis, requiring triangulation between methods (e.g. see SQW et al., 2015). Moreover, spillovers are relevant to this RTI policy, and these require some form of case-based research that seeks to track through how the original intervention has contributed to these effects. In essence, therefore, for most RTI policies, a single evaluation approach is unlikely to yield satisfactory findings.

For evaluators, developing a sound classification of these issues can require in-depth research with those involved in delivering the RTI policies. This participative approach should help evaluators to develop programme theories that better reflect the realities of policies, and ultimately evaluation design that is more appropriate and sensitive to the characteristics of policies and how they are implemented.

Aspect	Simple version	Non-simple version	Examples from RTI policies of non-simple version
Nature of implementation and engagement with the policy	Businesses/ organisations benefit on an individual basis from the policy	Multiple partners are involved when businesses/ organisations engage with the RTI policy (Complicated)	Collaborative R&D schemes RTI infrastructure Demand-side measures
Simultaneous causal strands	Single (at least primary) causal strand	Multiple causal strands (Complicated)	Various RTI policies, for example: specific outcomes relating to an R&D project supported by a policy (e.g. progress through technology readiness), alongside other outcomes such as development of innovation capabilities and feeding back into the research base
Alternative causal strands	Experience of the policy is the same/ similar, with broadly the same causal mechanism	Different causal mechanisms depending on context (Complicated – where variation can be categorised/ coded; Complex – where experience essentially bespoke)	RTI infrastructure: businesses' experience can vary, for instance as they select the support that meets their needs; feedback loops may also result in refining existing or bringing about new R&D projects
Timescales to outcomes	Same/similar for those benefiting from the policy	Variation in timescales to outcomes, e.g. reflecting technologies and markets (Complicated)	Collaborative R&D schemes and RTI infrastructure: the timescales to commercial benefits for businesses potentially vary from under 5 years to 15/20+ years
Policy objectives	Single, e.g. focussed on economic	Multiple, e.g. combination of economic, societal and system (Complicated)	Demand-side measures, where the purpose is to bring about economic growth, domain-specific objectives (e.g. clean energy) and change within the innovation system
Non-linearity and disproportionate outcomes	Linear causality and proportional impact	Feedback loops and the potential for a critical tipping point to bring about a large ultimate effect (Complex)	Demand-side measures, where small initial effects (e.g. increased initial take-up) can lead to a large ultimate effect (e.g. through feedback to innovators and wider diffusion)
Emergent outcomes	Outcomes can be pre-identified, e.g. increased R&D spend and business performance metrics of those directly involved	Outcomes dependent on the interactions between organisations, and how the behaviours are influenced (Complex)	Demand-side measures where effects rely on system changes, such as the interaction between different organisations to create appropriate frameworks

Table 1: Complicated and complex aspects of RTI policies // Source: Author, drawing on Rogers (2008)

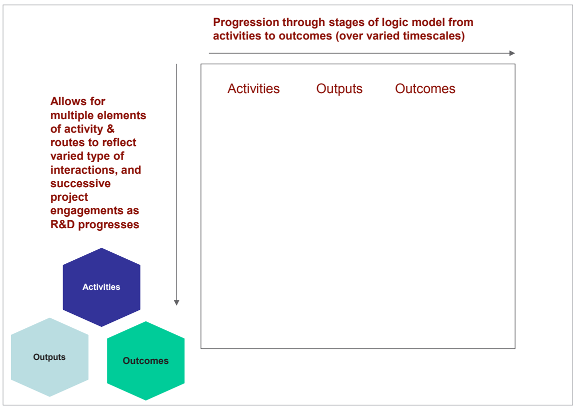
There is a risk that the logic models and programme theories for interventions with complicated and/or complex characteristics become too "messy", with every box in the logic seemingly linked in some way to every other box (Rogers, 2008). An alternative is to adopt a common structure or framework, within which a series of bespoke theories and logic models (or sub-theories) can be developed for individual projects or 'cases' (Cook, 2016; SQW and Cambridge Econometrics, unpublished). These sub-theories are particularly appropriate where there are numerous alternative causal routes, e.g. because an intervention is highly tailored in its implementation to particular contexts. This approach also facilitates building evidence on recursive feedback loops and interactions that are emergent, because it is designed to be easily refined and additive. Figure 1 provides a stylised example, illustrating three steps that show how more complicated aspects can be incorporated and how the approach can be tailored to different contexts:

- the overall structure is shown in step 1
- the population for a relatively simple intervention is shown in step 2
- and how this might be used to consider alternative or simultaneous causal strands and also new interactions/activities is shown in step 3.

FIGURE 1: USING FLEXIBLE PROGRAMME THEORIES TO FACILITATE EVALUATION

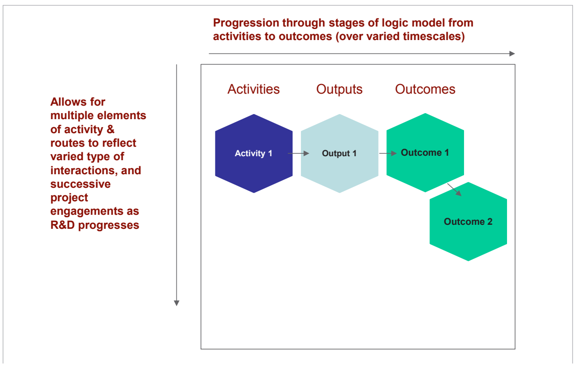
Step 1

All activities, outputs, and outcomes are allocated an individual 'tile' to be included on the logic model 'slate'. The 'slate' has two axes: a horizontal axis setting out the progression from activities through to outcomes (for which an underpinning theory of change is developed), and a vertical axis, that enables multiple forms of support or outputs/outcomes to be captured within the logic model.



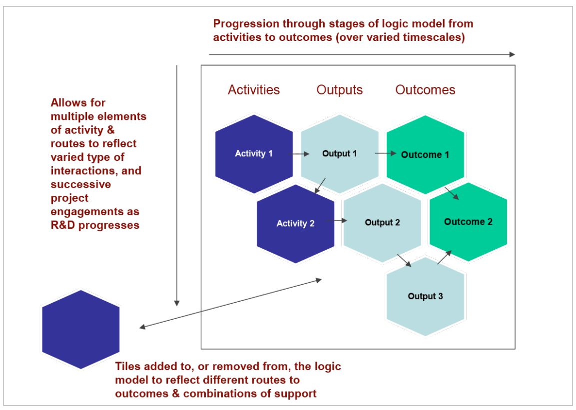
Step 2

The logic model 'slate' is populated with the relevant activity/activities, with the expected outputs and outcomes flowing from these identified and included – and a postulated theory of change developed. The approach opposite sets out a relatively simple generic logic model where one activity is provided, leading on to a single output and two outcomes



Step 3

The logic model 'slate' is updated through the addition of more 'tiles' based on evaluation evidence (e.g. if further activities, output, outcome combinations are identified) or to reflect increasingly complex and complicated routes to outcomes. The example opposite sets out a theory of change where two activities delivered in parallel lead to a range of outputs and outcomes.



Source: Author, drawing on SQW and Cambridge Econometrics (unpublished)

USING A THEORY-BASED APPROACH

As indicated in Table 1, most of the RTI policies examined here have complicated and/or complex characteristics, which pose challenges to evaluation. Responding to this, our work to scope evaluation approaches has frequently drawn on these determinants in recommending mixed methods within an overarching theory-based approach (e.g. SQW and CE, 2016). For example, contribution analysis is one such approach that can be taken, i.e. examining whether there is strong evidence that the intervention, rather than other factors, was critical in causing the outcomes observed (distinct from evaluating what would have happened in absence of the intervention). In the case of RTI policies such as collaborative R&D schemes and investments in RTI infrastructure, this is likely to involve collating a range of evidence in order to test, from different perspectives, the contribution of the policy under examination. This may include evidence from:

- project-specific case studies and beneficiary interviews to test, bottom-up, the contribution of the intervention to outcomes
- interviews/case studies with indirect beneficiaries to test the extent to which spillovers may have been achieved, and how far these relate back to the original intervention
- technology mapping combined with interviews with sector experts to assess, from a top-down perspective, the contribution that an intervention has made to more systemic change or technology development.

Such approaches can provide, in a transparent way, an assessment of whether, how and in what context, RTI policies have brought about their intended outcomes and also unintended outcomes. However, the extent to which the outcomes can be quantified and monetised will be limited, even at the level of individual beneficiaries of policies, let alone at the level of the policy overall. This may leave unanswered the policy-maker's question relating to the value of the policy. In some cases, a partial assessment may be possible here. Again, relating back to the classification of the characteristics in Table 1, for RTI policies or for the aspects of RTI policies that are merely (!) complicated, and for which a single or small number of key outcomes can be observed or assessed, a quasi-experimental approach could be used. Therefore, for at least part of the policy, some value can be ascertained. There is an important communications issue here, which relates to the tendency for audiences of evaluations to focus on what can be counted. Given the potential for RTI policies to lead to spillover effects and disproportionate outcomes that cannot be quantified, there are risks that these receive insufficient attention, thereby under-stating the effects of the policies, and creating perverse incentives for implementation to focus too much resource on the more direct routes to outcomes.

LEARNING POINTS

Several key lessons are relevant for evaluators, policy-makers and deliverers involved in RTI policies. First, the characteristics of interventions can be important determinants of evaluation approaches. However, classifying policies by these characteristics is not always neat and straightforward, because policies may often have combinations of simple, complicated and complex features. Second, there is a need to develop appropriate programme theories and logic models, which particularly draw on the perspectives of those delivering on the ground. Third, for interventions with complicated or complex characteristics theory-based evaluation approaches provide an important option where counterfactual-based approaches are inappropriate or to complement counterfactual-based approaches. Finally, even where parts of RTI policies can be evaluated using counterfactual-based approaches, policy-makers need to be alert to the partial story provided and the potential for such results to distort behaviours.

REFERENCES

Cook, J. (2016) 'Making the infinite countable? Responding to the challenges when evaluating innovation policy', Presentation to the UK Evaluation Society Conference 2016

Cook, J., Macdonald, B. and Pates, R. (2013) 'The Pareto Principle: the importance of the vital few in business support programmes', SQW Insight Paper

Hof, B., Koopmans, C., Lieshout, R. and Wokke, F. (2012) 'Design of a Methodology to Evaluate the Direct and Indirect Economic and Social Benefits of Public Investments in Space', SEO Economic Research Technical Note

Jordan, G. (2010) 'A Theory-Based Logic Model for Innovation Policy and Evaluation', Research and Evaluation, Vol.19 No.4

Rogers, P. (2007) 'Theory-based evaluation: Reflections ten years on', New Directions for Evaluation 114: 63-81

Rogers, P. (2008) 'Using Programme Theory to Evaluate Complicated and Complex Aspects of Interventions', Evaluation, Vol. 14 No. 1

SQW, Cambridge Econometrics and BMG Research (2015) 'Evaluation of Smart', Report to Innovate UK

SQW and Cambridge Econometrics (2016) 'The Aerospace Technology Institute: Scoping study to establish baselines, monitoring systems and evaluation methodologies', BIS Research Paper No. 271

SQW and Cambridge Econometrics (unpublished) 'High Value Manufacturing Catapult Impact Evaluation Framework', A Report to the HVM Catapult

Weiss, C.H. (2000) 'Which links in which theories shall we evaluate?' New Directions for Evaluation 87: 34-45

White, H. (2009), 'Theory-based impact evaluation: Principles and practice', Journal of Development Effectiveness 1(3): 271-84

White, H. and Phillips, D. (2012) 'Addressing attribution of cause and effect in small n impact evaluations', 3ie working paper, June 2012

AUTHOR

JONATHAN COOK
Director at SQW Ltd
SQW, 1st Floor 43 Chalton Street, London NW1 1JD,
United Kingdom
Phone number: +44 (0) 20 7391 4105 or +44 (0) 7715 071501
Email: jcook@sqw.co.uk

EVALUATION OF THE IMPACT OF R&D SUBSIDIES USING A MATCHING APPROACH

MARUSCA DE CASTRIS

The paper provide some empirical evidence of the impact of Italian Technological Fund for innovation. The policy measure was designed to promote the introduction of innovation or to stimulate research activity. R&D public subsidies data by firm are linked to panel data for Italian firms on sales, fixed assets, value added, employment. The impact of incentives in different group of firms is stressed. Counterfactual analysis is introduced to get significant estimates of the effects of R&D subsidies on treated firms. MDID estimators is used. The results are controversial as in previous studies.

INTRODUCTION

Both academic scholars and policy makers are debating the effectiveness of incentive system that boost firms' competition enhancing innovation and research and development (R&D) efforts. In the last 10 years, the objectives of the Lisbon Strategy (the objective of increasing R&D expenditure to 3% of GDP) have accelerated the growth rate of public R&D support but the sign and the size of the effects on firms' R&D expenditure and performances is an open question. Spurred by the increasing share of public resources devoted to supporting innovation activity, a growing body of literature has investigated the effectiveness of R&D subsidies. The findings are mixed and controversial. David et al. (2000) revise the results of forty years of empirical studies and find that there is no conclusive evidence in favour of public support. The unconvincing empirical results could mainly be explained by the difficulties in isolating the impact of innovation subsidies from the confounding effects induced by other factors. In particular, participation in these programs is generally endogenous and the selection bias is pervasive. Economists and econometricians deal with the problem of inferring the effect of a policy by using different evaluation methods (Blundell-Dias, 2009).

The study analyzes the effect of public R&D subsidies on firms performance and innovative efforts using a counterfactual approach based on a non-experimental method. The main concern is to assess the effectiveness of public R&D support on firm's performances analyzing whether the sign and the size of the effects depend on the size of the firms and on its technological level.

The Fund for Technological Innovation (FTI). The study compares subsidized firms with not subsidized ones using a counterfactual approach based on a MDID (Matching Difference-in Differences) estimator. The empirical analysis is carried on a detailed and informative database including companies awarded at least one R&D grant. We have data on the size of subsidies, from

the administrative archive, and balance sheet data from Bureau Van Dijk database. We estimate the impact of the subsidies on revenues, material and immaterial investment, value added, employment, labour productivity and profitability.

The empirical literature that evaluates the impact of public R&D on measures of performance is scarce and the results are not unique if we exclude expenditure on R&D.

LAW 46/1982: THE FUND FOR TECHNOLOGICAL INNOVATION

Among R&D subsidies to firms, law 46/1982 is one of the most relevant law to promote private investment in the field of research and innovation in Italy. The law creates two instruments to found R&D and innovation: the Fund for Research Credit and one that regards specifically the institution of a Fund for Technological Innovation (FTI).

The instrument operate following two ways: a direct subsidy to investment and an indirect subsidy for subsidized credit.

The selection procedure of the benefited firms is carried out by the Ministry of Industry. Firms apply demand and project and, through a procedure of enquiry, the competent office of the Ministry ascertain which firms satisfy the conditions required to get financial support.

The procedure makes use of a penalties when firms do not respect the programme interrupting the funding and forcing them to return the received amounts. The procedure does not consider the risk of non-additionality that is the hypothesis in which firms would have carried out the project in any case, also in the absence of public incentives.

THE IDENTIFICATION STRATEGY

L. 46/82 uses a planned selection process because subsidies are assigned to projects, and so to firms, following policy's criteria. This means that treated and not treated firms are different respect to their structural and financial dimension. Only a randomized assignment of subsidies could ensure that the two groups are not different. We are conscious that the selection system produces some types of selection bias that certainly influence the average outcome of treated and not treated firms. For example, larger firms characterized by high profit and capital intensive may achieve better results also in the absence of subsidy. Moreover, the possibility of being subsidized increases if the firm has better relationship with banks, has an effective management and the project is clear and well structured. Each factor can influence firm performance. For these reasons, the evaluation strategy aims to decrease the selection bias associated with a firm's observable and not observable characteristics.

The main observable characteristics which affect selection bias, are the factors considered more important to be eligible by the policy makers. For innovation project, economic sector and firm size can be relevant in the selection mechanism. EU rules assure higher incentives share to SMEs because the low size reduces the likelihood of access to credit.

In order to control for these effects, in the analysis we utilize information on firm size (measured by the number of employees).

Management ability and inclination to innovate are the major not observable characteristics.

We assume that other local factors are constant over time, and the effect can be captured by a firm fixed effect. In this set, we also include other not observable variables affecting the decision to participate, such as the quality of firm management and its propensity to risk, the quality of the R&D produced by the firm and productivity effects related to the geographical location of the firms, which are only partially captured by the previous covariates. These factors are all intrinsically related to each firm, and can be considered invariant over the analysed period.

THE EVALUATION MODEL

To identify the impact of L.46/82 using a matching technique we need that the control group satisfy two main conditions: (a) before the policy, the control group is very similar to the treated group (b) the control group is a very good control for the selection process.

We assume that the time dimension (the time when firm presents the project) and the space dimension (regions) are not relevant in respect to the selection problem. Under this hypothesis (which we verify below with several robustness checks) we pool projects across different regions. In this way, an overlapping area of firms with the same propensity to be subsidized (they are in both the treated group and the control group) is available and a matching estimator is a feasible instrument to determine the and the control group) is available and a matching estimator is a feasible instrument to determine the effects of L. 46/82.

The matching estimator assumes that selection can be explained purely in terms of observable characteristics. In this case the conditional independence assumption (CIA) holds, it means that the outcomes of non treated units are independent from the participation status conditioned to the observables. The consequence of CIA is that for each subsidized unit, observations of not subsidized unit on outcome variable with the same covariates realization constitute the correct counterfactual.

The ability of matching to reproduce an experimental framework depends on the availability of the counterfactual. Hence, the second matching assumption is that all treated units have a counterpart in the not treated population and any one constitutes a possible participant. The main advantage offered by the matching method is that it does not require any assumption on the functional form of the dependency between the outcome variable and the observed covariates. On the other hand, if there are a high number of covariates, it may be difficult to identify a not subsidized firm to match with every subsidized firm, unless the sample is huge. This obstacle is overcome with the Propensity Score Matching (Rosenbaum and Rubin, 1983). The correct use of a propensity score also requires that firms with the same propensity score must have the same distribution of observable (and non observable) characteristics independent to the treatment status.

This hypothesis is called the "balancing hypothesis" and can be tested using the approach presented in Becker and Ichino (2002).

In the case of L. 46/82, the weak unconfoundedness (CIA) hypothesis is theoretically not satisfied because we do not know the selection procedure. To implement the matching technique, we define the treatment group as the set of firms subsidized by L. 46/82 and the control group is made up of the rejected applicant firms. The outcome variable (calculated as compound annual growth rate) of interest is the performance, profitability and employment indices; the covariates refers to observed firms' characteristics such as size, activity sector and research cost.

Differences between subsidized and not subsidized outcomes persist also after conditioning on observables; in our analysis different regional or time fixed effects can affect the outcomes. We can correct for this potential cause of selection bias supposing that differences across regions are considered constant over time (Bernini, Pellegrini, 2011). The hypothesis is tested using a robustness analysis. Under this assumption a possible strategy to correctly evaluate the impact of L. 46/82 is to combine Matching with a DID estimator (MDID).

MDID consider first-difference outcomes on a pre-program period, in order to remove selection on time-invariant unobservables, both for subsidized units and unsubsidized ones, the latter selected

through a matching method, and compare the different outcomes to remove selection on observables (Smith and Todd, 2005; Blundell and Costa Dias, 2009). The MDID weakens the identifying assumption for matching by allowing non-observed time-invariant variables to influence performance (Bryson et al., 2002).

As usual, three statistical assumptions guarantee the validity of Matching and MDID estimation. The first assumption regards the Stable Unit Treatment Value Assumption (SUTVA), which requires the program not to have any effects on non-participants. The second assumption, regarding only the MDID, is the conditional independence of increments, that is, in the absence of the program, average variations of pre-program outcomes are identical among treated and untreated firms. Another hypothesis considers that the change occurred in the period before–after the treatment is the same for control firms and treated ones, regarding the observable component of the model and the non observable time trend. The assumption is rational if the treated firms have common characteristics with the non treated ones. After all, the assumption of common support requires that for each treated unit of the program there be observationally identical untreated units.

The effect of the treatment on the treated firms can be estimated over the common support of the covariates using the matching diff-in-diffs estimator (Blundell and Costa Dias, 2009).

RESULTS

The first step to estimate the impact of the policy is the specification of the propensity score model. We adopt a Logit specification of the treatment dummy variable (T), which is equal to one if firm has received the subsidy and zero otherwise. For the identification of covariates, we consider variables on fixed assets, sales, labour cost. Size is also controlled with dummies for medium or small firms. Localization is controlled with a dummy on the southern regions. The adopted specification also reflects that the selection procedure is not linearly based on the three main indicators and the interaction between the main indicators and dimension is introduced. The ratio labour cost and turnover per capita at time zero is used to control for pre-program firm productivity, approximating unobserved management ability.

ATT is estimated using the MDID technique, implemented by a Stratification matching estimator. The presence of some anomalous data (as signaled by the large difference between median and mean across indicators) indicates a need to trim the subsidized and the not subsidized firm samples at the 5 and 95 percentiles. We impose the common support restriction in all the estimations in order to improve the quality of the matches.

L. 46/82 has a significant positive effects on total fixed assets, employment and Research and advertising cost of the sample of subsidized firms.

In general, the study doesn't find significant positive effects on turnover, intangible assets and productivity. This highlights the absence of additionality of the subsidy. The positive effect on employment can be regarded as the increasing demand of high skilled workers employed in R&D activities, especially to design the proposal project.

Only medium firms gain the advantage of the subsidy as shown by return on investment, while the large firms can realize their project also in the absence of the incentives.

The evaluation of the effects of FTI subsidies on manufacturing sector shows a positive impact of subsidy on fixed assets and on research and advertising cost. This suggest that firms invest to increase

capital accumulation more than they would do in absence of the incentive. Unfortunately these investment do not produce significant effects on employment and firm performance.

The northern and central regions show better results than the whole country; the impact is significant positive on employment, turnover, fixed assets and ROI. This effect depend on the different territorial distribution of innovative Italian firms.

CONCLUSIONS

This article provides new evidence on the impact of public R&D funds; it highlights some positive effects still not came out of previous studies. It analyzes if the participation to FTI program leads on average to higher performance at the firm level. By means of a nonparametric approach, we compare the outcome of subsidized firms to a matched control group of not subsidized ones. The analysis of the effectiveness of the R&D subsidy is carried out using a counterfactual approach: treated firms are matched with control firms for each investigated aspect. The selection of control group is very careful in order to guarantee the closest (reliable) likeness to treated firms.

The information collected in our dataset covers administrative data and balance sheet data for the time before the investment and for the time following the investment.

This has allowed for a deepen analysis of the casual effect of public R&D subsidies. The casual effect identified is significantly positive for employment while it is significantly negative on productivity.

The conclusion of the study is still ambiguous: we have some issues to deal with to achieve a more comprehensive result. First of all, to improve the propensity score estimation controlling for more covariates able to differentiate treated and non treated firms, in order to reduce the selection bias effect. In this way, it could be useful to get information about firm's previous experiences in the field of technological innovation and R&D activities. Second, R&D investment can be influenced by the neighbouring innovation firms that can set barriers to entry and to get high skill workers.

REFERENCES

Aerts, K. and Schmidt T. (2008) Two for the price of one? Additionality effects of R&D subsidies: A comparison between Flanders and Germany, Research Policy, 37 (5): 806-822.

Almus, M. and Czarnitzki D. (2003) The effects of public R&D subsidies on firms innovation activities: The case of Eastern Germany, Journal of Business and Economic Statistics, 21: 226–236.

Czarnitzki, D. and Toole A.A. (2007b) Business R&D and the interplay of R&D subsidies and product market uncertainty, Review of Industrial Organization, (31)3: 169-181.

David, P. and Hall, B., (2000), Heart of Darkness: Modelling the Public-Private Funding Interactions Inside the R&D Black Box, Research Policy, 29, 1165-1183.

Rosenbaum, P., Rubin, D.B., 1983. The central role of the propensity score in observational studies for causal effects. Biometrika 70, 41–55.

AUTHOR

MARUSCA DE CASTRIS
Department of Political Science, University of Rome "Roma Tre"
Via Gabriello Chiabrera, 199, 00145 Roma

e-mail: marusca.decastris@uniroma3.it
Mobile +393334542575
phone +390657335344

EVALUATING PUBLIC SUPPORTS TO THE INVESTMENT ACTIVITIES OF BUSINESS FIRMS: A META-REGRESSION ANALYSIS OF ITALIAN STUDIES

ANNALISA CALOFFI, MARCO MARIANI AND ALESSANDRO STERLACCHINI

The use of public funding to foster different types of private investment is a common practice in many countries. For those belonging to the EU, the European Commission has established, since long time, specific guidelines in order to avoid that national (and regional) supports to business companies hamper competition. Moreover for the current programming period of the European Structural Funds, 2014-2020, the Commission requires an ex-post evaluation, based upon counterfactual methods, of the policy measures providing financial aids to private firms. Such an obligation to scrutinise how this portion of tax-payers money is spent has probably raised the cheers of some Italian experts and opinion leaders: finally, there will be the chance to proof that, in Italy, public incentives to business firms are most of the times a waste of money and, thus, should be drastically reduced (Giavazzi et al., 2012).

Actually, the foes of public incentives to enterprises do not know or guiltily neglect that in our country, over the last decade, the number of evaluations concerned with this topic was remarkable. In fact, considering the time span that goes from 2003 to 2015, we found 43 published studies on the effectiveness of public incentives to the investment activities of Italian firms. It must be stressed that this number is confined to the micro-evaluation analyses carried out in compliance with the methodological standards of the so-called "econometrics of programme evaluation" (Imbens and Wooldridge, 2009) and, as such, fully satisfy the above mentioned requirement of the European Commission.

Only a scant minority of these empirical works reports negative effects or no effect at all of the provided incentives, a finding that clashes with the liberalist vulgate invoking a retreat from public supports to private firms. However, as stressed by Stanley (2008) a simple vote-counting of studies (by distinguishing those reporting "positive", "insignificant" or "negative" effects) could be misleading. First of all, "statistically significant results are often treated more favourably by researches, reviewers and/or editors; hence, larger, more significant effects are over-represented. [...] Without some correction for publication bias, a literature that appears to contain a large empirical effect offers little, if any, reason for accepting this effect." (Stanley, 2008, p. 104). Moreover, studies using larger samples of firms are likely to find more statistically significant results (either positive or negative) than those based on smaller samples (Card et al., 2010). Finally and most importantly, rather than simply establishing what is the prevailing effect, a more interesting question for both researchers and policy makers is whether there are some factors (such as the chosen estimation technique, the type of incentive, the targeted beneficiaries, etc.) that increase the probability of such an effect. In this respect, literature reviews could provide useful insights. However, each survey contains a degree of subjectivity because the reviewer chooses the studies to be included and, although she tries to be as much comprehensive as possible, she attaches different weights to the selected works in order to identify the reasons of why contrasting findings are likely to emerge.

The approach that attempts to consistently address all the above issues is that of Meta-Regression Analysis (MRA), that is a "regression analysis of regression analyses" (cf. Stanley and Jarrell, 1989, p. 299). Being based on a quantitative exam of the literature, MRA allows one to test whether there are publication biases, as well as if the results change with the model specification and estimation method. Moreover, from a policy perspective, the approach is useful to identify whether the change or the probability of a given outcome (e.g. increase of investment expenditures, improvement of firm performance, etc.) is affected by some features of the policy measure (e.g. types of incentives, eligible beneficiaries, public bodies managing the intervention, etc.).

Garcia-Quevedo (2004), Negassi and Sattin (2014), Castellacci and Mee Lie (2015), Gaillard-Ladinska et al. (2015) apply this method for analysing the effects of public incentives on the R&D activities of business firms, while Kluve (2010) and Card et al. (2010) perform a MRA for some active labour market policies implemented, respectively, in Europe and worldwide.

In this paper we apply a MRA to the already mentioned empirical studies that have estimated the effects of public support to the investment activities of Italian firms. To our knowledge, this is the first application of a MRA to such type of micro-evaluation studies. In order to achieve a sufficiently high number of observations we have considered 43 published works, providing about 470 estimates, concerned with the impacts of different public incentives (subsidies, soft loans, tax credits, public loan guarantees) on different kinds of outcomes (inter alia, expenditures on R&D as well as other categories of tangible and intangible investment, innovation activities, debt consolidation, firm performance in terms of employments, sales or productivity). Because of the wide spectrum of outcome variables taken into consideration, we perform a MRA by using as dependent variable a binary indicator equal to one when the public support has generated a "significantly positive" result. Thus, as in Garcia-Quevedo (2004), Kluve (2010) and Card et al. (2010) the analysis refers to the sign and statistical significance of the policy effects.

In addition to these previous works, our meta-regression model for the probability of a positive, and statistically significant, treatment effect also includes a term of unobserved heterogeneity at the study level. In particular, we estimate a random-intercept multilevel model that allows us to take into account for the possible correlation between treatment effect estimates stemming from a same article. The reasons for such correlation could be due to the unobserved ability of the authors in framing the study or obtaining credible estimates, or also it might depend on their determination to search for particular results.

We find that the occurrence of positive effects is not affected by the number of firms considered in the empirical analysis as well as by whether the study was published in a journal: accordingly, there is no reason to believe that publication bias affects our estimates. The most striking finding of our meta-analysis is that a positive effect of the policy is more likely to emerge when the measured outcome is directly targeted or immediately affected by the policy. Indeed, depending on the type of programme, the occurrence of positive treatment effects increases when the outcome variables refer to R&D expenditures or R&D employees, amount of capital investment, receipt of favourable bank loans or lower interest rates, rather than to other indicators of firm performance. This finding is not surprising. In fact, although effects on the latter type of outcomes are often hoped for by policymakers, they may emerge only after a rather uncertain chain of events, which is difficult to assess.

With respect to some common policy schemes, our findings show their probability of success is non-negligible. If there exist any differential in probability of success between the government levels that may deliver the programmes, this differential is favourable to regional governments. As a possible explanation for this result, it can be argued that regional policymakers, being particularly aware of the specific features and behaviour of local firms, are able to design and implement more effective policy measures than their national counterparts. In addition to that, however, is should be recalled that the studies on regional programmes considered in our analysis mostly refer to northern and central Italian regions, which, according to European standards, enjoy a decent quality of government and administration (see Rodríguez-Pose and Garcilazo, 2015).

REFERENCES

Card, D., Kluve, J., Weber, A. (2010), Active labour market policy evaluations: A meta-analysis, Economic Journal, 120 (548), F452-F477.

Castellacci, F., Mee Lie, C. (2015), Do the effects of R&D tax credits vary across industries? A meta-regression analysis, Research Policy, 44 (4), 819-832.

Gaillard-Ladinska, E., Non, E.M., Straathof, B. (2015), More R&D with tax incentives? A meta-analysis, CPB (Netherlands Bureau of Economic policy Analysis), Discussion Paper 309.

García-Quevedo, J. (2004), Do public subsidies complement business R&D? A meta-analysis of the econometric evidence, Kyklos, 57 (1), 87-102.

Giavazzi, F., D'Alberti, M., Moliterni, A., Polo, A., Schivardi, F. (2012), Analisi e raccomandazioni sul tema dei contributi pubblici alle imprese. Rapporto al Presidente del Consiglio e Ministro dell'economia e delle finanze e al Ministro dello sviluppo, delle infrastrutture e dei trasporti redatto su incarico del Consiglio dei Ministri, 30 aprile 2012, 23 giugno 2012.

Imbens, G.W., Wooldridge, J.M. (2009), Recent developments in the econometrics of program evaluation, Journal of Economic Literature, 47 (1), 5-86.

Kluve, J. (2010), The effectiveness of European active labor market programs, Labour Economics, 17 (6), 904-918.

Negassi, S., Sattin, J.F. (2014), Evaluation of public R&D policy: A meta-regression analysis, University of Delaware, Department of Economics Working paper NO. 2014-09.

Rodríguez-Pose, A., Garcilazo, E. (2015), Quality of government and the returns of investment: Examining the impact of cohesion expenditure in European regions, Regional Studies, 49 (8), 1274-1290.

Stanley, T.D., Jarrel, S.B. (1989), Meta-regression analysis: A quantitative method of literature surveys, Journal of Economic Surveys, 19 (3), 299-2307.

Stanley, T.D. (2008), Meta-regression method for detecting and estimating empirical effects in presence of publication selection, Oxford Bulletin of Economics and Statistics, 70 (1), 103-127.

AUTHORS

ANNALISA CALOFFI
Department of Economics and Management, University of Padova,
Padova, Italy.
Email: annalisa.caloffi@unipd.it

MARCO MARIANI
IRPET – Tuscany's Regional Institute for Economic Planning,
Florence, Italy.
Email: marco.mariani@irpet.it
+390554591295

ALESSANDRO STERLACCHINI
Department of Economics and Social Sciences, Marche Polytechnic University
Ancona, Italy.
Email: a.sterlacchini@univpm.it

THE INDO-EUROPEAN COLLABORATION IN THE SCIENCE, TECHNOLOGY AND INNOVATION: LOOKING AT THE FRAMEWORK CONDITIONS AND THE OUTCOMES

TERESA DE OLIVEIRA, ANGIOLILLO SEAN, KASTURI MANDAL AND PARTHASARATHI BANERJEE

The attention to collaboration in research has grown significantly over the last decade indicating that collaboration plays an important role in the development of research. In fact even the role of international collaboration has been receiving growing attention. The present paper deals with international research collaboration, takes the case of the Indo-European cooperation and examines the concrete example of the participation of India within the context of the Seventh Framework Program and the successor, Horizon 2020. We will analyse two different layers linked to the international research collaboration; the external conditions of the funding schemes and the internal project level. The external perspective highlights those specific frameworks conditions applied to India and the very distinct outcomes in terms of India´s participation within the FP7 and Horizon 2020 that it has brought into light. At the internal project level, the perspective of the users of those programmes played significant roles in accomplishing international research collaboration. Furthermore, this paper addresses two main research questions: 1) how the frameworks conditions applied to India affected the participation of India within FP7 and H2020; 2) how the project coordinators benefitted from of the international research collaboration.

The chosen research approach is mid-range theory building up from aggregate empirical data, "aiming at integrating theory and empirical research to explain a specific set of phenomena" (Merton, 1968). The concept of international research collaboration is particularly relevant when applied to our study and it can be defined and operationalized in many different formats, including: researcher exchange; formal intergovernmental agreements on scientific cooperation; meetings and workshops; international large-scale facilities, collaborative projects, publications, international large-scale facilities and the establishment of laboratories (Georghiou, 1998).

In this paper, international research collaboration is seen to be a joint research activity within a common aim or shared objective (Katz and Martin; 1997; Shrum et a.2007), among scientists based at public research institutes in different countries and regions. Under this definition lies a conception of "deep collaboration involving a division of labour and creative contributions from all partners, rather than weaker forms of collaboration" (Laudel, 2011). Within this paper, we will be looking at the international research collaboration within the programmes whose raison d'être is to "foster global cooperation in research through project support" (Georghiou, 1998), especially the Seventh Framework Programme and its successor programme – Horizon 2020.

India and the European Union have strong ties and the collaboration in the field of STI has gained scope and increased importance, since the conclusion of the Science and Technology Cooperation Agreement (STCA) between India and the European Community in 2001. The Seventh Framework Program by the virtue of its structure and procedures made it more accessible to international collaboration and added new opportunities for Indian organisations, bothfrom academic and technological fields. India's participation in the EU Framework Programme has been growing tremendously: from 36 participating organisations in FP4 (1998-2002) and 39 in FP5 (1998-2002); 142 participants in FP6 (2002-2006) and more than doubled in FP7 (2007-2013) to 305 participants. India's leadership in projects also reached a peak in FP7, with Indian participants taking the coordinator role in 19 projects in FP7, up from 9 in FP6. India was the fourth most active Third Country in terms of participation (305 participants in 181 projects) and in terms of receiving financial contribution (Đ35.8m) from the European Commission — behind only Russia, the United States, and China.

When we examine the external factors that enabled India to participate so fully, the availability of EU funding was the decisive one. In reality, from the FP4 (1998-2002) until FP7 (2007-2014), India has been able to participate to the EU calls in a logic of automatic eligibility for EU funding. Under the EU Seventh Framework Programme, specific themes targeted cooperation with Third Countries and established a network of different stakeholders and thus trying to improve scientific collaboration and define priority areas. As suggested by Sonnenwald (2007), the framework conditions for practicing international research collaboration can be favoured by introducing mechanisms of funding research centres or offering funding for collaborative research. Also, the availability of help positively influences international research propensity (Birnholtz, 2007; Amabile et al., 2001; Sargent& Waters, 2004). The funding, the resources available, and the institutional support appeared to be one of the factors that can influence the most international research collaboration. In the case of the Indian participation within the FP7, the external factor of available funding had a very positive impact on India's participation. Conversely, the Horizon 2020, has introduced a new EU Approach to International STI Collaboration: Indian participants are no longer eligible for automatic funding in the classic collaborative projects. The rationale for the policy decision to remove applicants of BRIC+Mexico countries from automatic funding of collaborative projects can be found in an official European Commission communication document, "Enhancing and focusing EU international cooperation in research and innovation: A strategic approach". Its first chapter titled "A Changing World", the document notes Europe's continued leadership role in research and innovation in terms of expenditure, publications, and patents, but recognises the following change in the global STI landscape: "over the past decade, however, the landscape has evolved rapidly. Global research and innovation were, until recently, dominated by the European Union, the USA and Japan. As the emerging economies continue to strengthen their research and innovation systems, a multipolar system is developing in which countries such as Brazil, China, India and South Korea exert increasing influence. The changes in the global research and innovation landscape explain the following organisation of countries, where India is therefore seen to be more appropriately grouped with "Industrialised countries and emerging economies" rather than developing countries. This shift at the policy level and the external funding conditions, as stated Sonnenwald (2007), played in the first years of implementation of Horizon 2020 a negative impact in terms of the participation of India within the above-mentioned programme. Comparing the first two years of FP7 and nearly the first two years of Horizon 2020, this amounts to an approximately more than an 85% decrease in the number of Indian participants, the number of projects with Indian participants, and the total value or cost of those projects. Not surprisingly, according to our analysis, most, if not all, of the countries included in this policy change have seen negative consequences in terms of their overall participation in the programme. India though, seems to have been especially impacted.

There have been changes in the strategic goals of each generation of EU S&T programs. The FP6 had formulated its thrust on searching for and sustaining networks of research collaborations around sustaining nodes. This generation thus sustained oligarchy of nodes. It was assumed that such an oligarchy would remain operational and active even without the prop of support from program funding. Therefore the EU designed a strategic shift in the designing of the next program, the FP7. In this later program voluntarism replaced the core oligarchic nodes. The voluntarism of FP7 was founded on the programmatic belief that open calls for program funding would elicit responses from the markets of S&T actors, the individual persons and organizations across multiple countries, including international cooperation. Such responses from individual actors, just as free competition in a market place entails, would lead to emergence of bright and brilliant novelties and ideas of S&T. The S&T outcome being the global common property resources of such an open market place of voluntaristic individuals competing for program funding would be innovative too by virtue of its novelty. The voluntarism of FP7 thus replaced the node-actor bounded oligarchic structure of STI global production. The FP6 had used network measures as the STI indicators while the FP7 began using number of individuals and their distinct non-networked collaboration as the indicators of the STI outcome.

While, the Horizon 2020 raised its strategic objectives to the level of science driven innovation. The innovation for Horizon 2020 is understandably of larger import causing the restructuring of very large socio-technical systems. Necessarily, such great and onerous tasks of restructuring would remain outside the horizons and the competencies of non-state voluntary and small actors. The Horizon 2020 therefore began with a rapid switch over from voluntaristic individuals to the interplay of state as actors. This new and the current program of the EU involves coordination between the states as S&T parties, and this program makes a demand upon coordinating states to shape up its domestic voluntary as well oligarchic actors into a coordinated system based on equally basis. The evaluation of Horizon 2020 necessarily therefore involves a measure of inter-state S&T coordination supplanted and orchestrated by an intra-state coordinating capability and capacity.

Each of these EU STI programs took up clearly targeted and strategically unique set of actions supported by a large financing package. As a result, the STI indicators that could measure and thus evaluate a program of previous generation become inappropriate for the successor generation of program. In other words, we put forward a claim of this paper that STI indicators are dependent on the strategies of a STI program. Therefore, in order to undertake evaluation of a program, we need to look up the specific strategy and then design a corresponding set of STI indicators which can capture the degree of success that this program can attain over a specific period. Both FP7 and H2020 deploy the instrument of collaboration in particular international collaboration, even so with unique outcomes. In both the cases international collaboration nurture the growth of STI and evaluation of each program informs us differently on the differentiated pathways of the development of STI. Deeply connected with the framework conditions for forging and developing the international research collaboration are the outcomes and the benefits of that collaboration. The research question here is: How did the project coordinators benefit from of the international research collaboration?

Within the framework of set of semi-directed interviews, more than 20 projects coordinators of FP7 projects mentioned that one of the drivers to collaborate with EU was indeed the advancement of knowledge and academic excellence. In fact, as many studies on international research collaboration have found that researchers with a reputation for academic excellence tend to collaborate with other researchers across the world (Mcdowell andMelvin 1983; Piette and Ross 1992; Rijnsoever and Hessels2010; Vafeas 2010). The contents analysis of the interviews shows that, in general, there was a positive response from the interviewees. The general conclusions are that on the Indian side there was, a positive response from the project coordinators and members of the consortium being associated with EU related projects as this gave them a good exposure to EU based science institutions and researchers; on the European side there was, in general, a positive experience for the European research teams and the possibility to interact with significant research challenges in the fields of Water, Energy and Health. In conclusion, scientific and cultural experience seem to be one of the most significant features of EU – Indian cooperation projects under the framework of FP7.

Interestingly enough, in the case of our sample of interviews, the advancement of knowledge - proved to be a very positive outcome of the projects for both sides. Thus, Indian scientist participants stated that projects were not so much about advancing a piece of frontier knowledge but stressed how the project enabled them to address a problem in the Indian context. From the European side, it was mentioned that the fact of partnering with India brought complementarity and a better understanding in some specific challenges, notably in the field of Environment, Health, etc. Overall, in the FP7 project context, advancement of knowledge often allowed the acquiring new perspectives on joint problems.

Faced with this twin tasks this paper brings out an analysis of external conditions and the internal perspective of the users within FP7. Moreover, and since an evaluation is based largely on the assessment of international collaboration in S&T and RTI, the exercise of evaluation implicitly proposes a set of potential indicators of S&T and RTI. In other words the current paper strives to place the exercise of evaluation on the foundation of international collaboration and thus various phases of collaboration we would like to argue, indicates both the performance and the generational shifts across programs spread over decades. Our empirical results also therefore suggest financial and attentional resources, such as, external funding conditions, advancement of knowledge as internal features, played significant roles in accomplishing international collaboration between India and Europe.

BIBLIOGRAPHY

Beaver, D. D. (2001) 'Reflections on scientific collaboration, (and its study): Past, present, and future', Scientometrics, 52: 365–77.

Bozeman, B., Fay, D. and Slade, C. P. (2013) 'Research collaboration in universities and academic entrepreneurship: The state- of-the-art', Journal of Technology Transfer, 38: 1–67. and **Rogers, J. D.** (2002) 'A churn model of scientific knowledge value: Internet researchers as a knowledge value collective', Research Policy, 31: 769–94.

Boga, M., Wells, S., Abbey, S., Bergman, A., Chandler, V., Embleton, L. (1992) Advancing social work practice in the health field: a collaborative research partnership. Health and Social Work 17: 223–235.

Georghiou, L. (1998) 'Global cooperation in research', Research Policy, 27: 611–26

Laband, D. N. and Tollison, R. D. (2000) 'Intellectual collaboration', Journal of Political Economy, 108: 632–62.

Laudel, G. (2001) 'Collaboration, creativity and rewards: Why and how scientists collaborate', International Journal of Technology Management, 22: 762–81

Merton, R. (1968) Social Theory and Social Structure. New York: Free Press

Katz, J. S. and Martin, B. R. (1997) 'What is research collaboration?', Research Policy, 26: 1–18.

Kreiner, K. and Schultz, M. (1993) 'Informal collaboration in R&D: The formation of networks across organizations', Organization Studies, 14: 189–209.

Ulnicane, Inga, Why do international research collaborations last? Virtuous circle of feedback loops, continuity and renewal, Science and Public Policy 42 (2015) pp. 433–447 doi:10.1093/scipol/scu060

AUTHORS

TERESA DE OLIVEIRA
researcher and project manager at the Centre for Social Innovation, scientific Institute based in Vienna, Austria.

ANGIOLILLO SEAN
Head of Research Projects, IndoGenius, Private Company based India.

KASTURI MANDAL
Scientist, National Institute of Science, Technology and Development Studies (NISTADS), Indian Council of Scientific and Industrial Research

PARTHASARATHI BANERJEE
Goa Institute of Management, Goa, India

INTERNATIONAL RESEARCH LINKS OF EU13 COUNTRIES AND THE CONSEQUENCES FOR EU RESEARCH PROJECT PARTICIPATION: FP7 PARTICIPATION DYNAMICS AND THE PREREQUISITES TO EU RESEARCH FUNDING SUCCESS

NICHOLAS HARRAP AND MATHIEU DOUSSINEAU

The countries that joined the EU in 2004 and after (EU13) have low participation in EU research programmes such as the Seventh Framework Programme (FP7) and the current Horizon 2020 programme (European Commission, JRC-IPTS, 2015). These EU13 countries only accounted for 4.2% of the funding and 8% of participations for FP7. However, it has been claimed that this low level of involvement does not fully reflect their capabilities and potential (EU-12 Member States, 2011). Stated reasons include problems of the national research landscape, a lack of competitive funding environment and size and resources of these countries that means they do not have the capacity to compete in all research areas (European Commission, 2010).

Other work has studied the collaboration characteristics and how these affect the participation in framework programme (FP) funded projects. With regard to firms, it has been demonstrated that those that are not already connected struggle to have a central position in a FP network (Autant-Bernard et al., 2007). At the same time studies on the 5th and 6th FPs have shown that regions that co-publish frequently do not receive a disproportionate share of funding (Hoekman et al., 2013). The same study showed that the effect of funding on co-publication activity is especially significant for regional pairs that did not intensively co-publish before participation. It suggests also that the returns to FP funding are highest when involving scientifically lagging regions.

Participation in FP7 for EU13 countries deteriorated when compared to FP6 (European Commission, JRC-IPTS, 2015). Therefore, it is pertinent to consider how strong the links are inside and outside FP7 networks to see if EU13 were at a disadvantage to participate or whether it is a mechanism whereby they can increase their integration into the international research ecosystem. Policy questions could be asked: how to better network to improve participation to EU research funding?; how to better deploy the links created by EU research programmes to strengthen participation in international science generally?; or how to break the cycle of lack of international research links? A lack of such links means fewer opportunities to participate in EU research programmes and so few opportunities to increase international research links.

Cross border research collaboration and networks are not the only factors affecting the EU13 level of FP7 participation. However, understanding the characteristics of networks could help in designing strategies to overcome deficiencies, including by the use of a combination of different funding sources. Such synergies are currently the focus of a strategic push through the implementation of smart specialisation strategies for the combination of ESIF funds and Horizon 2020. The objective was therefore to better understand the research links of EU13 countries and the policy implications and understand their importance for project participation and whether project participation enhances international links.

Collaboration is defined by the activities (research and observation, experimentation, data collection, publication) undertaken by researchers who are working together on common research projects (Wagner et al, 2002). Collaborations can materialize by a research contacts that can link researchers and lead to co-publications and/or patent co-invention. Therefore, collaborative activity indicates that researchers are members of networks and/or are developing their networks through collaborative activity. While this work takes collaborative research networks as the analytical focus it should be noted that networks are wider than this and collaborative research networks are a subset of the overall network structure.

With regard to research organisations, the integration between institutions has increased over time, due to the involvement of institutions within multiple projects (Pohoryles 2002, Barber et al 2006). This indicates a move towards a more integrated European Research Area (ERA) and that collaboration within European funding frameworks has led to more durable links between collaboration partners. Furthermore, there has also been a significant tendency for the same institutions to participate in consecutive FPs with recurring collaboration between the same organisations within the FPs (Roediger-Schluga & Barber, 2006).

While on the one hand these phenomena can be seen as an increased integration of the ERA, they could, on the other hand, be seen as leading to the domination by a core of actors (Breschi and Cusmano, 2004) to the detriment of those outside the core. These core actors can effectively be seen as forming a 'club'. If these actors are concentrated in certain countries then these countries are also, by default, the dominant core countries.

AIMS AND OBJECTIVES

One factor for low participation from EU13 countries that has been hypothesised is that the EU research programmes are dominated by the old EU15 Member States and the new Member States are effectively locked out of these "clubs" of tight networks. While some studies have demonstrated the importance of being connected (Autant-Bernard et al., 2007) others have indicated that FP participation has less to do with network effects such as the creation of closed clubs and more to do with institutional characteristics such as size and reputation (Lepori et al., 2015). Others were unable to find evidence of country cluster regarding FP7 participation (Rauch and Sommer-Ulrich, 2012).

Therefore the aim of this work is to understand the structure and characteristics of collaboration networks in different settings and investigate whether the collaboration domain of FP7 is different to the international collaboration domain in general.

Therefore, a first objective is to understand the strength of links between different EU countries and the characteristics of their research networks in FP7 and the wider research community including those that are not necessarily participating in EU research programmes.

A second objective is to study the collaboration network characteristics to understand whether the differences between the domains can indicate whether countries are locked out of FP7 through a "club" effect by being unable to access the tight networks of the EU15 countries or whether the comparison indicates a more fundamental problem for EU 13 countries to access competitive funding (such as reputation of organisations, resources and governance).

Based on these aims and objectives three research questions were set that guide the work described in this abstract.

- What are the strongest links for each country in each domain: FP7 and co-author?
- What network characteristics are apparent for each domain: certain countries in small world networks/clusters?
- How are the collaboration and network characteristics for the EU countries different or similar within each domain?

METHODOLOGY

The FP7 contracts and proposals database 2007-2013 is used in order to match collaboration patterns with Bibliometric indicators (Scopus database of research output).

In order to investigate the existence of a 'club' effect maintaining some EU countries partially out of the ERA, we focus first on internal composition of FP7 consortia. A penetration rate of countries in FP project is calculated and matched with other types of indicators coming from the FP or outside (Bibliometric, patents) in order to see how this rate is linked to the capacity to influence the building of consortia from co-optation behaviour. This is a demonstration of the 'club' effect. The number of participants coming from the same country/region in a same project is also a sign of co-optation behaviour.

Salton's index is used to characterize the link between partners in research projects and co-authors in publications; it is a relational indicator of the strength of co-authorship links (Glänzel, 2001). In the case of publications, it is calculated by taking the total number of joint publications between two countries and dividing it by the square root of the total number of publications of the two countries (Glänzel, 2001).

While the number of publications can be considered a proxy for research activity so co-authorship of publications can be considered a proxy for collaboration. International collaboration is indicated when the authors' affiliation addresses are in two or more different countries.

The co-authorship can be attributed to a country (or region) through whole counting where every country with a contributing author is counted (Leydesdorff and Wagner, 2008). In fractional counting the country is assigned a fraction of the paper. For this study, where the link the important factor, whole counting is the most appropriate approach.

It should be noted that co-participation analysis of FP7 projects could measure a collaboration that will also be registered as a co-authored publication. To reduce any impact the analysis is undertaken with those publications citing funding from the EU being considered while acknowledging that not all publications will cite the funding source.

OUTCOMES

The outcome of this study is a contribution to a better understanding of the barriers affecting EU13 participation in large collaborative research programmes. This work contributes to the understanding of whether there exists a 'club' effect of big players in Europe hindering the participation of organisation based in New Member States. Furthermore, it offers an insight to the benefits of participation for EU13 countries on their international scientific engagement.

Further work could consider broadening the scope to take account of more applied networks research leading to patents through co-invention links and networks using patent data from the Patstat database.

REFERENCES

Autant-Bernard, C., Billand, P., Frachisse, D., Massard, N., (2007), "Social distance versus spatial distance in R&D cooperation: Empirical evidence from European collaboration choices in micro and nanotechnologies", Pap. Reg. Sci. 86, 495–519.

Barber, M., A. Krueger, T. Krueger and T. Roediger-Schluga (2006), "The network of EU-funded collaborative R&D projects", Physical Review E, 73

Breschi, S. and Cusmano, L. (2004), "Unveiling the texture of a European research area: emergence of oligarchic networks under EU framework programmes", International Journal of Technology Management, 27, 747-772.

European Commission, JRC-IPTS (2015), "Stairway to Excellence Facts and Figures: Poland", available at: http://s3platform.jrc.ec.europa.eu/documents/20182/117536/S2E_PL_national_profile.pdf

EU-12 Member States (2011), "Common Position Paper of the EU-12 Member States for the next Framework Programme", available at: http://ec.europa.eu/research/horizon2020/pdf/contributions/prior/common_position_paper_of_the_eu-12_member_states.pdf

European Commission (2010), "Interim Evaluation of the Seventh Framework Programme – Report of the Expert Group", available at: https://ec.europa.eu/research/evaluations/pdf/archive/other_reports_studies_and_documents/fp7_interim_evaluation_expert_group_report.pdf

Glänzel, W (2001), "National characteristics in international scientific co-authorship relations" Scientometrics, 51, 69–115

Hoekman, J., Scherngell, T., Frenken, K., Tijssen, R., (2013), "Acquisition of European research funds and its effect on international scientific collaboration", J. Econ. Geogr. 13, 23–52.

Lepori, B., Veglio, V., Heller-Schuh, B., Scherngell, T., Barber, M., (2015), "Participations to European Framework Programs of higher education institutions and their association with organizational characteristics", Scientometrics, 105, 2149–2178.

Luukkonen, T., Tijssen, R., Persson, O., Sivertsen, G., (1993), " The measurement of international scientific collaboration", Scientometrics, 27, 1.

Pohoryles, Ronald J. (2002), "The Making of the European Research Area.a View from Research Networks Innovation", Vol. 15, No. 4, 2002.

Rauch, M., Sommer-Ulrich, J., (2012), "Participation of the Central and Eastern EU Member States in FP7":Analysis, Evaluation, Recommendation – Summary, Fraunhofer Moez.

Roediger-Schluga, Thomas and Barber, Michael J. (2006), "The structure of R&D collaboration networks in the European Framework Programmes", UNU-MERIT.

AUTHORS

NICHOLAS HARRAP
European Commission - Joint Research Centre
Institute for Prospective Technological Studies (IPTS)
Edificio Expo
Calle Inca Garcilaso, 3, 41092 Seville
Spain
+34 95 448 83 62

MATHIEU DOUSSINEAU
European Commission - Joint Research Centre
Institute for Prospective Technological Studies (IPTS)
Edificio Expo
Calle Inca Garcilaso, 3, 41092 Seville
Spain

EVALUATING PARTICIPATION OF TOP CLASS UNIVERSITIES IN EUROPEAN RESEARCH PROGRAMMES: WHAT INSIGHTS FOR POLICY DEBATE?

EMANUELA REALE AND EMILIA PRIMERI

RATIONALE/ PURPOSE OF THE PAPER

Universities represent the major share of EUFPs participants, and their participation is highly concentrated in a small number of European research Universities. These universities usually are at the core of well established and successful international networks (Lepori et al., 2014; Reale et al., 2008), which often rank top at the European and international level as for their scientific reputation, scientific outputs, collaborations and PhD students training (Altbach, 2009). So far, EU funding seems to strengthen the concentration of resources in few countries and producing cumulative and self-reinforcement phenomena (Geuna, 2001; Lepori et al., 2014; Primeri and Reale, 2014).

Concentration of University participation in few organizations located in few countries raises question about to what extent EUFPs are a mean for European integration toward the ERA, or rather they are likely to reinforce the existing imbalances between countries as to the capability to enter networks and to develop research collaborations. In this respect, several evaluation approaches about the achievement of EUFPs programmes objectives and results have been developed, which generally focus on the capacity of EUFPs to foster large EU research policy objectives such as widening participation, improving integration, increasing synergies across member states and enhancing performance.

This paper instead wants to assess motivations and effects produced by the participation in EUFP at the level of Higher Education Institutions (HEIs); the analysis focuses on University participation to EUFP7, taking top-class research universities as specific cases.

The questions we address are: what indicators can provide better insights about top class universities participation to EUFPs? To what extent they can provide policy makers with useful information about Programmes design and implementation? We assume that motivations and effects of EUFPs participation are highly diversified also across top class universities, which mirror differences in EU programmes involvement of national governments and characteristics of national R&D systems.

The paper combines different methods, mixing both qualitative and quantitative approaches, and control the motivations and impact of EUFP7 participation in top-research universities with those of other European research universities.

THEORETICAL FRAMEWORK

A relevant development in evaluation purposes is the shift from a focus based on national aggregates to individual actors indicators, such as funding agencies or Higher Education Institutions (Lepori and Reale, 2012). Also literature underlines as indicators have expanded their role in understanding research programmes aims and objectives, rationales and major assumptions, beyond allowing measuring its impact and outputs, often shifting the unit of analysis from the programmes to the participants level. One example is the understanding and the evaluation of universities participation to EUFPs, which is generally based on indicators concerning data and statistics about selection, participation, and results (Lepori and Reale, 2012). More specifically indicators mostly concern a) institutions performance and funding, b) tools of participation, considering then thematic priorities addressed and main instruments used, c) collaborations, focusing on networks relationships, among academic institutions and between these and firms. Nevertheless EUFPs are policy instruments that contribute to academics organisational, cultural, and cognitive changes (Primeri and Reale, 2012); they are shaped through a complex political process of negotiations between motivations, interests and expectations of different stakeholders (Primeri and Reale, 2012; Lascoumes & Le Galès, 2005). In this respect, recent trends in evaluation underlines the need to move from measuring to learning purposes (Molas Gallart and Davies, 2006) thus from a linear evaluation model based on impact measurement to a formative role of evaluation aimed at supporting and boosting policy debates and learning from previous experiences.

Literature has often assumed that top class universities have a limited interest toward EUFPs, because of their applied nature, the administrative requirements and burdening rules of participation, and are mostly concentrated in national opportunities of funding, where they often have a preferential access (Henriques et al, 2009).

On the other side, several studies have underlined as the status of being "excellent" affect the participation and performance in the European Framework programmes (Nokkala, Heller-Schuh, Paier, 2011), and that excellent universities often represent the lion share of EUFPs participation and funding, mostly engaging as coordinator instead of partner in collaborative projects, being part of well established networks (of excellent partners) and concentrating the most in frontier research (Primeri and Reale, 2012).

In this paper we consider the Framework Programmes as a set of opportunities intended by the policy makers and then provided to the potential beneficiaries in the programme design; the mentioned opportunities are differently perceived and mobilized by the beneficiaries - the universities- on the basis of the research priorities and strategic aims of the institutions themselves (Reale et al., 2014). Inconsistencies between intended and provided opportunities supply evidences of possible shortcomings of the policy action; inconsistencies between provided and perceived opportunities shed light on problems related to the programmes design; differences affecting the perceived and mobilized opportunities provide evidences about different internationalization strategies. Different paths can therefore emerge in the engagement of Universities in EUFPs that are linked to the modes in which universities understand the opportunities supplied by the programmes and decide to act accordingly (Reale et al., 2014).

METHODOLOGY

The participation of HEIs to EUFPs was analysed to depict:

- Characteristics of participation across the specific programs, thematic priorities, and the likelihood of top universities to become coordinators than non-top universities;
- Motivations and lasting effects of participation that can differentiate them compared to other non-top universities. So far, the analysis would allow highlighting specific motivations and rationales for participation and impact produced for this sample of academic institutions. Also it would allow discussing the extent to which EUFP, and in particular EUFP7 with its focus on excellence and competitiveness, meet expectations of top universities (LERU, 2010);
- Contribution of participation in EUFPs to the University standing within European and non-European universities.

The empirical base of this work consists in 25 case studies on European high performing research universities participating in EUFP7. The sample selection is based on the number of participations in the several generation of FPs (from IV to VII), and the reputational standing of the organization in terms of research activities. More precisely the criteria adopted for the selection are the following ones (ranked by importance): a) world standing in research activities (positioning in the University Ranking – check done on Shanghai Ranking, Leiden Ranking, and Multi-rank rankings); b) high participation in EU FPs (baseline EU FP6 and FP7); c) a balanced presence of generalist universities and technical universities; d) Geographical representation, of different countries within EU28 in order to avoid a concentration of cases in few countries (Table 1).

Table 1 – Universities analysed through the case studies

University	Country
University Of Cambridge	UK
Politecnico Di Milano	IT
Univ. Bologna	IT
Aarhus Universitet	DK
The University Of Edinburgh	UK
Universidad Politecnica De Madrid	ES
Chalmers Tekniska Hoegskola	NL
University Of Uppsala	SE
Katholieke Universiteit Leuven	BE
The University Of Leeds	UK
The University Of Manchester	UK
King's College London	UK
Ecole Polytechnique Federale De Lausanne	CH
Karlsruher Institut Fuer Technologie	DE
Technische Universiteit Delft	NL
Technische Universiteit Eindhoven	NL
University College London	UK
Eidgenoessische Technische Hochschule Zuerich	CH
Karolinska Institutet	SE
University Copenhagen	DK
Universiteit Utrecht	DE
Danmarks Tekniske Universitet	DK
Lunds Universitet	SE
Pierre Marie Curies	FR
The University Of Nottingham	UK

The methodological approach integrates results from:

- Descriptive analysis on university participation and success rate based on E-CORDA dataset;
- Descriptive analysis based on the positioning of the 25 universities of the sample in the selected rankings (Shanghai, and Leiden Rankings), plus information on the universities from the ETER dataset;
- Qualitative information coming from a survey based on a common questionnaire developed on 100 universities participating in EUFPs (from EUFPIV to EUFPVII), of which 25 are those investigated through the case studies;
- Qualitative information coming from 30 in depth interviews to 25 top-research universities (in some cases more than one interview has been realized in the same University) selected among those addressed by the survey. The interviews content was analysed using the software ATLAS T.I, according to the dimensions listed in Table 2.

Table 2 – Dimensions for the analysis of the case studies

EUFPs main opportunities	Key dimensions assessed	Main items/indicators
Funding	Amount and duration of funding	Access to funding Consistency of funding with research objectives
Networking	Collaborations building and duration	Entering new networks Long lasting collaborations Public-private collaborations
Innovation and excellence	Building reputation	Improvements of reputation and excellence Achievements in terms of scientific outcomes Gaining leadership
	Knowledge improvement	Risky and innovative research Frontier research Access to new knowledge and equipment Knowledge acquisition, use of new equipment
	Knowledge mobilization	Training opportunities for researchers/Phd Researchers career
	Research outcomes	Development of intellectual property Dissemination of research outcomes
New fields	Fields innovation in research	Interdisciplinary Cross field research
Additionality	Added value of EUFPs	Benefits of EUFPs compared to national resources Possibility to replace EUFPs

The analysis of participation to EU research programmes mixes both qualitative and quantitative approaches in the evaluation design, allows capturing on the one side data about participation (e.g. number of projects awarded, funding, number of collaboration). On the other side, it allows framing EU programmes as policy instruments that can produce highly diversified effects with respect to the actors involved, more specifically between top research universities and other universities in Europe.

Finally differences in Universities participation to EUFPs would highlight the need to discuss about the balance between equity and excellence in the European Framework Programmes design (Arrow, 1993; Guri, 1986).

The main limitation of the study is that it investigates the government level of the HEIs and not the different researchers participating in the EUFPs. This limitation does not allow to depict differences between research fields, and to investigate individual motivation and behaviours of the beneficiaries. However the results provide interesting insights on how the opportunities provided by EUFPs have been then mobilized at institutional level.

FINDINGS

The study confirms that strategies and motivation -which drive top class universities participation to EUFPs, are different from those of other universities.

Three results emerge distinctively for top class universities from the survey. Firstly the possibility EUFPs provide to access funding to basic and high-risk research and the fact that they allow, especially through ERC projects, to carry out frontier research.

Secondly, answers highlight that EUFPs provide opportunities to strengthen top class universities existing positions in excellence networks, gaining further leadership positions in emerging networks, improving high-quality research and productivity, and training young researchers. Both are reported as perceived opportunities and as mobilised ones, underling as the initial expectations and motivations for participation are generally fully met. Both the mentioned items do not have the same importance for the non-top research universities, whose main motivations are linked to achieving additional funding and improving the international standing joining new networks.

Thirdly the shift and the substitution capacity of EUFPs compared to national research programs are concerned, questioning whether national scientific goals and priorities should be better designed keeping or not an overlap with the European ones. In this respect results highlight that there is not a “substitution” effect of European programs. Interestingly enough, the survey provides evidence of differences between the motivations (perceived opportunities) of beneficiaries and the benefits they effectively recognized as linked to the EUFP participation (Table 3).

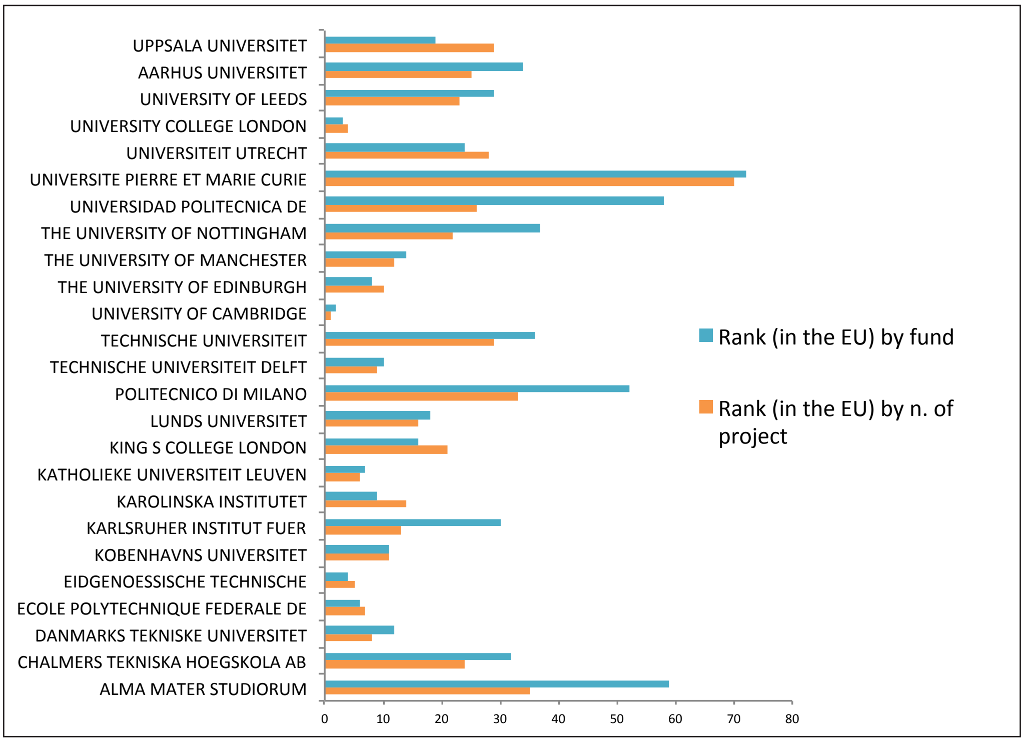
Table 3 – Man results from the survey

Motivations	**All universities**	**Top universities**
Funding	High	Not relevant
Reputation	High	Very High
R&D Outputs (excellence and productivity)	High	Relevant
Multidisciplinary research	Low	High
Training PhD/early researchers	Low	Low
Networking	Relevant	Very High
Leadership	Low	Very High
Benefits	**All universities**	**Top universities**
Funding	High	Relevant
Reputation	Relevant	Very High
R&D Outputs (excellence and productivity)	High	Relevant
Multidisciplinary research	Low	High
Training PhD/early researchers	Low	High
Networking	Relevant	Very High
Leadership	Low	Very high

EUFPs are considered unique means compared to other existing schemes at national and supra-national level. The uniqueness derives from the capability to join (and to lead) networks as well as to reinforce collaboration with the most reputed scholars in the different fields working in EU. Funding becomes a relevant item also for top research universities, as well as training because of the very positive appreciation of the Marie Curie actions. The mentioned items are assessed in a very different way in the other universities surveyed.

However, important differences across the top research universities emerge. Figure 1 shows the strong differences existing between the HEIs in the sample as to the participation in EUFP7.

Figure 1 – Ranking of the 25 selected universities by funding and project granted in EUFP7



Source: eCORDA 2015; Note: the lower the value the higher is the position of the University in the ranking

The reasons provided by the interviews confirm that top-class universities are not an homogeneous group of performers as to all the dimensions analysed (Tab. 2); differences in HEIs participation are mostly related to the availability of funding for research and development and on the academic standing of the universities. In fact, participation in EUFPs and positioning of the HEIs in the international rankings are highly correlated. Moreover, several organizational issues emerge as important elements of differentiation across top universities. Looking at a limited number of 5 world-class universities included in the case studies, we can see that several obstacles to participation (risk and cost of submitting a proposal, bureaucracy, confidence in the EU evaluation process and selection procedures) are not considered very important, while they have been mentioned in the other interviews as elements which can discourage scholars to participation, looking for other sources of funding. Another important item is the presence of a formalized ad hoc strategy for participation in EUFP, with the possibility of scholars that want to apply to have dedicated grants for the preparation of the project proposal.

Finally, the recognition of EUFP as a mean related to the ERA concerns (integration, in-ward and out-ward mobility) do not emerge as perceived and mobilized opportunities. Europe is an arena for competition and collaboration between research actors, and EUFPs are efficient means to consolidate high standing positioning of actors. Interestingly enough, also the mentioned item was declined in very different ways by the universities under examination; the very high-standing organizations present quite homogeneous views with no interest nor perception of ERA as an issu related to EUFPs, while the other top research universities articulate the answers according to different conditions existing at national level.

FINAL REMARKS

Some general observations can be drawn: the strategic importance and role of ERC and Marie Curie for excellent universities, which question the capacity of EUFPs of widening academic scientific quality. On this respect the programmes design should be also questioned, discussing whether a top down approach should be balanced with a more bottom up definitions of priorities and thematic areas to be addressed by Universities. Also results seem to confirm that excellence represents often the main objective to be achieved by EUFPs, which reinforce the participation ad success of high ranked academic institutions at the expense of equity and enlargement. This seems also to highlight the increasing influence of a managerial paradigm on European Framework Programmes (Young, 2015) with an emphasis of European policies on the importance of pursuing and measuring excellence and fostering competitiveness (EC-CREST, 2009).

To conclude the analysis of top universities participation in EUFPs provide interesting insights on drivers of universities participation and the capacity of the EU research programmes to meet needs and expectations of a broader set of universities instead of a narrow bulk of best performers, letting emerge more lock in mechanisms in EU research policy instruments design than tools for aligning and widening participation. The approach based on mixing quantitative and qualitative empirical evidences under the 'opportunity framework' confirms its capability to provide relevant insights for the implementation of the policy instruments.

ACKNOWLEDGMENTS

The authors acknowledge support from the European Commission through the study on "Universities engagement. An analysis of the role and engagement of universities with regard to participation in Framework Programmes", under the Framework Contract: 2012/s 144-240132

REFERENCES

Altbach et al. (2009): Trends in Global Higher Education: Tracking an Academic Revolution. A Report Prepared for the UNESCO 2009 World Conference on Higher Education

Arrow, K. J. (1993): Excellence and Equity in Higher Education. Education Economics, 1(1), 5–12

EC (2016): An analysis of the role and engagement of universities with regard to participation in the Framework Programme. Final Report. EC, Luxembourg

EC (2009) Mutual learning on approaches to improve the excellence of research in universities, CREST Fourth OMC Working Group, Final report. EC, Luxembourg

Enders, J. (2004): "Higher education, internationalization and the nation-state: recent developments and challenges to governance theory" Higher Education, 47, pp. 365-370

Geuna A. (1998): "Determinants of university participation in EU-funded R&D cooperative projects" Research Policy, 26, 677–687

Geuna A. (2001): "The changing rationale for European University research funding: are there negative unintended consequences" Journal of Economic Issues, 35, 607-632

Gornitzka, A. (1999): "Governmental policies and organisational changes in higher education" Higher Education, 38, 5-31

Guri, S. (1986). Equality and excellence in higher education — Is it possible? Higher Education, 15(1), 59–71

Hazelkorn, E. (2008): "Learning to live with league tables and ranking: The experience of institutional leaders" Higher Education Policy, 21(2), 193–215

Henriques, L., Schoen, A., Pontikakis, D. (2009): Europe's top research universities in FP6: scope and drivers of participation, EC JRC

Young, M. (2015): "Shifting Policy Narratives in Horizon 2020", Journal of Contemporary European Research, 11 (1): 16-30

Lascoumes, P.& Le Galès, P. (2005): Gouverner par les instruments, Les Presses de Sciences Po, Paris, France

Lepori, B.,et al (2014): Understanding factors influencing participation to European program of Higher Education Institutions. Paper submitted to the 19th International Conference on Science and Technology Indicators, Leiden, 3-5 September 2014

Lepori,B., Reale, E. (2012): S&T indicators as a tool for formative evaluation of research programs, Evaluation, 18(4) 451–465

Molas-Gallart,J. And Davies, A (2006): Toward theory-led evaluation - the experience of European science, technology and innovation policies, American journal of evaluation, Vol: 27 - pags 64 - 82

Neave, G. (1998): "The Evaluative State Reconsidered" European Journal of Education, 33(3), 265 - 284

Nokkala, T., Heller-Schuh, B., & Paier, M. (2011): "Ranking Lists and European Framework Programmes: Does University Status Matter For Performance In Framework Programmes?" In P. Teixeira, & D. Dill (Eds.), "Public Vices, Private Virtues? Assessing The Effects Of Marketization In Higher Education", pp. 111-139,Rotterdam: Sense Publishers

Primeri E., Reale E. (2012): "How Europe Impacts on Academic Research: The Transformative Potential of the European Framework Programmes" In: Effects of Higher Education Reforms: Change Dynamics, pp. 97-120. Rotterdam: Sense Publishers

Reale E., Nedeva M., Thomas D., Primeri E. (2014): "Evaluation through impact: a different viewpoint" Fteval Journal for Research and Technology Policy Evaluation, 39, pp. 36-41

Reale, E., ed. (2008): La valutazione della ricerca pubblica. Una analisi della Valutazione Triennale della Ricerca“, AA.VV, a cura di Emanuela Reale, Franco Angeli, Roma

AUTHORS

EMANUELA REALE
Corresponding author
IRCRES CNR
Via dei Taurini 19, 00185 Roma
Italy
e-mail: emanuela.reale@ircres.cnr.it
phone +390649937853

EMILIA PRIMERI
IRCRES CNR
Via dei Taurini 19, 00185 Roma
Italy

CRITERIA FOR SOCIALLY RESPONSIBLE RESEARCH PROCESSES – MAKING A DIFFERENCE TO RESEARCH IMPACTS?

JOHANNA FERRETTI AND KATHARINA HELMING

The debates about research's role in the further development of societies focus on research outputs and research's thematic orientation for tackling sustainability issues. Still, a scientific discussion emerges on the question of how research should be conducted to best contribute to solving societal challenges. The hypothesis stands that the design of the research process itself determines its outcomes and impacts in regard to sustainable development. This process is coined as socially responsible research. We identified a set of eight criteria which characterise socially responsible research processes with help of a literature review and numerous stakeholder workshops. In this paper we present the criteria identified and we report about results from test applications at various levels of research the research cycle. On this basis, we lay out a concept for prospective impact assessment of the set of criteria in order to test the hypothesis. We also seek for case study collaborations (with an agricultural focus) at international level for further testing of the set of criteria.

INTRODUCTION

Research closely interacts with society. The activities and findings of actors in the science system have multiple impacts in and on society. Two developments can be recognized in the recently renewed debate about the role of research in society. First, science is increasingly promoted for economic growth and job creation and considered as a central pillar of knowledge-based economies (Leydesdorff 2010). Research and the private sector need to collaborate from early on and ensure knowledge transfer, to drive innovations and thus contribute to prosperity and functioning of industrialised countries (European Commission 2011). Another type of development stresses research's contribution to sustainable development and tackling the grand societal challenges such as efficient energy supply or sustainable agriculture. In this context research has a strong thematic alignment which for instance takes shape in the so called sustainability research (e.g. Wiek, Ness et al. 2012). These two streams of debate partially overlap and are oftentimes discussed in combination (OECD 2010). Both call for science and research to take over more responsibility in the development of societies.

Second, and parallel to this the outputs produced by researchers and science systems are examined more carefully, particularly in the form of so called research impact assessments. Countries such as the UK (cf. Penfield, Baker et al. 2013) or Australia (Jones, Castle-Clarke et al. 2013) have begun to assess the quality and impacts of research, to base decisions for allocation funding on it, and to showcase "the value of taxpayers' investment in university research" (Group of Eight and Australian Technology Network of Universities 2012). The German Expert Commission on Research and Innovation commissioned an evaluation of the government's two science Pacts for universities and non-university research organisations (Möller 2016). Although some of these assessment exercises also capture the wider benefits of research, most of them consider impacts primarily in form of bibliometric and exploitable results such as patents (Penfield, Baker et al. 2013, Weißhuhn and Helming 2015).

The above shows that the demands of science's role in society mainly focus on economically exploitable outputs and with respect to a sustainable development on the "what" question of research. On the other side of the coin, research assessment practices strongly focus on research outcomes. In both developments, the process dimension of research, the "how" question, has not been central. Departing from the assumption that the design of the research process is a key factor for shaping scientific knowledge production and research outcomes to take over social responsibility and contribute to sustainable development we developed a set of criteria which defines socially responsible research processes and offers a framework for reflection.

DEVELOPMENT OF A CRITERIA SET TO DEFINE RESPONSIBLE RESEARCH PROCESSES

The set of criteria supports the alignment of research processes against the leitbild of "socially responsible research". The criteria partially concern established approaches which have been subject to different levels of analysis and which have been institutionalised to varying degrees at the different stages of the research cycle (Ferretti, Daedlow et al. in preparation). A systematic compilation of such criteria did not exist so far though. The set of criteria contributes to closing this gap by combining the single criteria in one systematic framework.

Eight criteria were identified in a comprehensive literature review and in iterative expert workshops and comprise: ethics, integrative approach, interdisciplinarity, transdisciplinarity, user orientation, reflection of impacts, transparency, and dealing with complexity and uncertainty. Based on the review and the expert discussions, so called "fact sheets" were prepared which describe the relevance, contents as well as approaches for reflection for each criterion. The eight criteria together with the fact sheets represent the framework for reflection (Daedlow, Podhora et al. submitted). The criteria can be subsumed under the question "How is research done?" and "For/ with whom is research done?". They claim to be applicable to all types of research (basic to applied research) and all scientific disciplines and should ideally be considered in all six phases of the research cycle (in reference to Schnell, Hill et al. 2008), from strategic agenda planning, carrying out research projects to evaluation and monitoring as shown in Figure 1. In their interaction and when reflected in research processes, they produce outcomes facilitating sustainable development.

Generally, the framework was assessed as relevant for the reflection of research processes in social responsibility by the test "runners". The compilation of criteria – partially considered as intuitive knowledge – was appreciated because allowing for a structured approach to socially responsible research processes. However, incentives were necessary in order to facilitate reflection in research activities. In this context trade-offs in the application of the set of criteria were addressed, particularly with efficiency, competitiveness, and freedom of research. The test runs moreover highlighted the need for further exploration of the potential implications of the set of criteria at different stages of research.

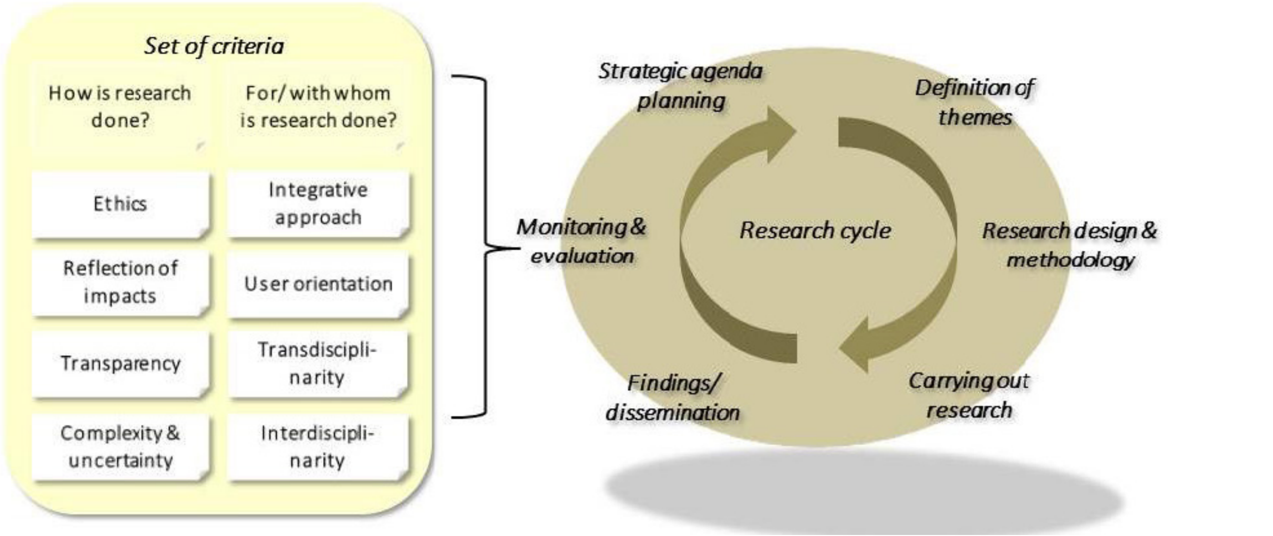


Figure 1: The eight criteria in the research cycle

TEST APPLICATIONS OF THE SET OF CRITERIA

We conducted test runs of the set of criteria to first-time examine in how far it is suited to reflect research processes regarding their social responsibility. The test runs took place in three institutions of the Leibniz Association (one of the three non-university research organisations in Germany), the Academy for Spatial Research and Planning, Leibniz Institute for Zoo and Wildlife Research (IZW), and the Leibniz Institute of Ecological Urban and Regional Development[1].

Researchers of different hierarchy levels (project management, science management) involved in projects and activities (joint projects, PhDs, basic and applied research) were asked to consider their own research based on the reflection framework. Subsequently they were interviewed to provide insights on the suitability of the framework.

FINDINGS AND CONCEPT FOR AN IMPACT ASSESSMENT OF THE APPLICATION THE CRITERIA SET

The test runs demonstrated that implementation of the presented criteria for assessment of the social responsibility of research is ambitious, time consuming, and its advantages and trade-offs are not yet well defined. A careful implementation process would therefore include three activities: training, piloting, and impact assessment. A toolbox with training material and hands-on application support could help individual researchers and research organisations to implement a reflection process with the criteria set. This toolbox should cover means for enabling, implementing, monitoring and assessment. A piloting activity would accompany researchers and research organisations in their quest to applying the criteria. It would identify obstacles, challenges and opportunities for successful implementation and identify best practice examples. An impact assessment would be necessary to test the hypothesis that the approach indeed delivers an added

1 Further test runs will be conducted in May and June 2016 in health and energy production related institutions and projects as well as with representatives of the Federal Ministry of Education and Research

value and that this would outweigh possible trade-offs. A methodology needs to be developed to test the advantages, drawbacks and side-effects of implementing the set of criteria in research of different disciplines, thematic orientations, and time horizons.

For the follow-up project we take the set of eight criteria as our hypothesis, arguing that when applied as a framework of reflection, research processes are more likely to impact in direction of socially responsible research and thus sustainable development. Thereby we will pursue the following research questions

- What are the implications to research processes at different stages when reflecting the criteria set?
- How can each of the criteria be operationalised to be applied in research processes?
- Which benefits and trade-offs arise from such requirements?

To answer the research questions we will strain an international comparison to explore how other countries frame social responsibility in research processes and how they evaluate the impact of research activities. We will develop a method and identify indicators to „measure" implementation and the implications of reflecting the criteria. To test the method and indicators case studies will be conducted to explore the application of the set of criteria to research processes at different stages and different progress levels (ex-ante, ex-post, accompanying research). We particularly seek case studies from research in the field of agriculture or land use (e.g. food security, or soil and ecosystem services) since they allow to consider approaches from basic as well as applied research, to build on existing research impact assessment literature in this field (Gaunand, Hocdé et al. 2015), and to consider a sector which is just beginning to develop a common understanding of innovation processes (Bokelmann, Doernberg et al. 2012).

ACKNOWLEDGEMENT

This research was carried out in the context of the LeNa project ("Guidelines for a Sustainability Management in Non-University Research Organisations", www.lena-projekt.de), funded by the Federal Ministry of Education and Research.

REFERENCES

Bokelmann, W., et al. (2012). Sector study on the analysis of the innovation of German agriculture (Sektorstudie zur Untersuchung des Innovationssystems der deutschen Landwirtschaft) Berlin, Commissioned by the Federal Office for Agriculture and Food (Bundesanstalt für Landwirtschaft und Ernährung, BLE).

Daedlow, K., et al. (submitted). "Socially responsible research for sustainable development: a review and integrated framework for operationalization." Current Opinion in Environmental Sustainability.

AUTHORS

JOHANNA FERRETTI
Leibniz Centre for Agricultural Landscape Research (ZALF)
Eberswalder Straße 84, 15374 Müncheberg
Germany

e-mail: ferretti@zalf.de
Phone: +49 334 3282 4075

KATHARINA HELMING
Leibniz Centre for Agricultural Landscape Research (ZALF)
Eberswalder Straße 84, 15374 Müncheberg
Germany

RESPONSIBLE RESEARCH AND INNOVATION: IMPLICATIONS FOR RESEARCH EVALUATION AT UNIVERSITIES

EMANUELA REALE

RRI is an emerging discourse at national and European level for the governance of science, which includes public engagement, science education, gender dimension, and ethics, open science. RRI is targeted as a process devoted "to align research and innovation with the values, needs and expectations of society" (EC 2011), to produce a 'right impact', to make the motivations and the intentions for actions in research and innovation more democratic. In November 2014, the Rome Declaration addressed directly governments, research funding organizations and research performing institution to actions toward RRI. The discussion with stakeholder specifically pointed out the need to change the perspective of evaluation in order to understand how far RRI is progressing in research organizations (Reale, 2014)

The paper assumes that responsible research challenges research organizations, Universities first and foremost, and evaluation with new questions, which are related to the progress toward the assumption of the RRI dimensions in the university governance, and require new criteria and indicators. In fact, RRI cannot be assessed under a performance-based approach based on efficiency and effectiveness. RRI asks for reflexivity that universities and research communities should adopt as normal component of their research practice, about the ultimate goal of their efforts and the role they are playing in society.

We argue that research evaluation shall improve the formative approach to assess opportunities and characteristics of the stakeholders' engagement in research. It means that activity indicators, rather than performance indicators of actual implementation can provide a useful approach (Lepori and Reale, 2012). The university internal governance and the decision-making shall evolve toward including the new dimension of responsibility; evaluation can have a strong role, supporting the debate, providing evidences about results achieved and open challenges, feeding up learning processes and rethinking about research aims and directions.

THEORETICAL BACKGROUND

For a long time the metaphor of a tacit negotiated contract between science and society has been used for the empirical analysis; the contract foresaw a clear division of tasks between the two parties: government (as representative of the society) supplies money; the scientific community provides knowledge retaining the power to decide the research agenda, methods and tools to guarantee integrity and social benefits (Guston, 2000). The mentioned approach, often operationalized using the principal-agent theory, was then criticised by authors claiming the need to move toward a new conceptualization of the contractual relationships, where society is not only represented by governments, and science comprises a set of different actors contributing to the production of a "socially robust knowledge" (Gibbons et al., 1994; Martin, 2003). Hessel and colleagues (2009) proposed a framework to analyse the contract between science and society based on the delegation of different tasks. The contract impacts how science operates by the way of the relevance that scientists want to achieve; so changes in its components "can be analysed in terms of the credibility circle" based on the Latour's and Woolgar's conceptualization of scientists' struggle for reputation (Hessel et al., 2009 396).

RRI goes beyond the mentioned approaches, suggesting the need to improve the democracy in decision-making, the institutions' and scientists' responsiveness, and generate new spaces of public dialogue to question about choices of academic research and desired results (Rip and Joly, 2012). Several pieces of literature investigated the key RRI dimensions in science and society relationships, and the mechanisms able to explore the type of impact a decision of science might produce (Felt et al., 2007; Guston, 2012; Owen et al., 2012). Other more recent contributes pointed out governance principles and requirements for responsibilisation of research organizations to ensure the quality of interactions, positioning and orchestration, and developing supportive environments (Randles and Laredo, 2012; Kuhlman, 2016).

The quoted literature figures out two main problems. The first is the uncertainty of the results of research activities, which make difficult to understand today the future developments of knowledge, and to direct science and innovation toward specific desired results. The second problem, strictly related to the former, is the freedom of individual research activity and the autonomy of the research organizations.

PROBLEMS IN RRI EVALUATION

Looking at the actual implementation of responsible research at university level, there are signs that the responsibility issue is mainly conceived as part of the third mission activities, foremost those related to public engagement. For example, the National Coordination Centre for Public Engagement (NCCPE) in UK seeks to support a culture change in universities. Another example is the last guidelines of the Italian Agency of Evaluation of University and Research-ANVUR for the assessment of third mission (ANVUR, 2014). The two examples both follow a performance-based approach of evaluation, where indicators are suitable to produce metrics aimed at ranking and rating universities performances.

Nonetheless, RRI assumes a different perspective of how research shall be carried out and how innovation is generated, which needs i) to improve democracy in the decision making, ii) to improve the institutions' and scientists' awareness before society, iii) to open new spaces of public dialogues, and iv) to question about choices of academic research and desired results. Thus, evaluation shall focus on changes occurring to attitudes and behaviour in research practices and to opening of opportunities for learning at the organizational level. In this respect one can distinguish between evaluation of research and evaluation of RRI research and/or universities.

It is also important to recall that RRI is different from *public scrutiny* of science and *public understanding* of science, asking "new interactions in the risk society and with continuing trust in science even if specific developments may be criticized" (Rip, 2003 35). RRI also differentiates from *accountability*; accountability refers to transparency, efficiency and effectiveness using public money; RRI affects processes actors, and contents of the decision-making asking responsibility about the effects that can be produced and how they are desirable ones. On the other hand, so far the boundaries between RRI concept and *precautionary principle* are not so clear, the main difference being the defensive perspective of the former with respect to RRI, which ask for a more proactive engagement of science toward society.

Thus, the traditional role of research evaluation (accountability, strategic change, and decision support) is challenged and revised by the opening up of boundaries between science and society, and the new actors involved (Rip, 2003).

EVALUATING RESPONSIBLE RESEARCH AT UNIVERSITIES

RRI is likely to differently affect the university, so one can broadly distinguish between two main categories: strategies and planning at central and meso government level, and actual research project at laboratory level.[1] The former is supposed to prepare conditions (opportunities) and spaces for the implementation of responsible academic research (gender, public engagement, ethics, etc.) in the academic governance. The latter is supposed to incorporate knowledge and values coming from society into the research practices. Some problems affect both the mentioned levels. On the one hand the capability of universities to act as strategic actors is limited and subject to several constraints especially as to the control they can exert on knowledge production (Whitley, 2008); on the other hand, scholars have problems to disclosing the content of their research and the methods used, how the problems are managed and solved, and the new result they want to achieve. The limitations affect the extent to which universities are able to promoting their role in society; in this respect evaluation is a mean that could support the formation of an attitude toward responsible research and the execution of action in this context.

Following Stilgoe and colleagues (2013), supporting the development of RRI shall integrate several, often fragmented components, which can be summarized along four dimensions:

Anticipation, which is devoted to manage the social uncertainty through generating "abilities to bridge the cognitive map between present and future" (Barben et al., 2008, 991; Guston, 2014). It means to spread the use of techniques and methods like scenarios, technology assessment, planning that shall try to figure out possible answers to the "what if" question to anticipate (not predict) and to shape desirable futures;

- *Reflexivity*, based on institutional practices mirroring and minding the consequences, the system of values and the activities actually developed;
- *Inclusion*, making the non-academic people inside the decision-making process of the science;
- *Responsiveness*, as capability to change direction of research activities "in response to stakeholders and public values and changing circumstances" (1572).

In table 1 the quoted classification of the RRI dimensions in the different government levels inside universities are summarized. (Tab.1).

Tab. 1 – Classification of the RRI dimensions at university

	Anticipation	Inclusion	Reflexivity	Responsiveness
University Strategies and Planning	Open the future possibilities on emerging fields and technologies (e.g. foresight, risk assessment, scenarios)	Open up the decision making process to external voices (e.g. consensus conference, deliberative mapping, focus group)	Questioning about responsibility (code of conduct, guidelines, standards)	Questioning about transparency (e.g. open access, transparency, project design)
Actual Research projects	Integrating the results of anticipation in the research activities	Including the participation of non-academic stakeholders from the very beginning of the research projects	Building connections between internal values and external believes	Changing the directions of the activities under development when the knowledge and control on possible effects and control are insufficient

Activity indicators are one possible approach to assess the RRI dimensions; they are generally used to present an indeterminate progress state, providing information to decision-makers, scholars, and stakeholders, who is conducing the activity, what was done, and where it was working. Activity indicators indicates what feature a user is currently using; they are aimed to deepen how far the commitment of the institution, group or individual is going in the right direction in order to pursue its main goal. Furthermore, activity indicators should allow assessing tools and practices toward RRI in terms of (Callon et al., 2009):

Intensity (if the action toward RRI occurs at an early stage and how large it is in terms of actors involved and processes affected),

Openness (how diverse and varied is the group of actors involved),

Gravity (if the discussion is on actual items related to the future of the science).

1 Laboratory level in this scheme refers to the research groups' activities and construction of science (Latour and Woolgar, 1979).

Table 2 provides a representation of the type of information the activity indicators can supply, to assess RRI within the university planning and within the research actual research projects.

Tab. 2 – Activity indicators characterizing RRI

Intensity	When	Before starting the policy-research action (e.g. before deciding the strategic plan; before addressing new techniques that implies ethics problems) Alongside the implementation of the objective At the end of a period of time of implementation
	Size	One-time use vs continuous use External /internal people involved Adequacy to the objective pursued
Openness	Internal	Diversity of internal actors as to governance level, academic position, disciplinary sectors-areas
	External	Diversity of external actors (groups from different organizations, cultures and practices)
Gravity	Relationship	Relationships between the issues selected for RRI practices and the university strategy
	Importance	Importance of the issues selected for RRI practices for the university quality and competitiveness Importance of the issues selected for RRI practices for the society

The matrix in Table 3 is one example of combining the dimensions of RRI and the activity indicators for assessing –under a formative approach, how far the university is moving toward responsible research, for instance using one tool or a combination of tools (e.g. a code of conduct, public engagement events, gender equality rules, etc.).

Tab. 3 – Characterizing the progress toward RRI through tools and practices

	Intensity	**Openness**	**Gravity**
Anticipation	Activities (regular, one-time) linked to figure out future possibilities and emerging fields in the strategic planning for research	Is participation an issue at stake in anticipatory practices and what is its relevance in the final output	What are the effects produced by spaces implemented for anticipatory purposes on the university strategy
Inclusion	When and how far there are attempts to opening the decision making processes and what actors are involved	How far the debate about RRI issues involve internal actors and at what level	The contents elaborated during inclusive decision-making process improve the quality of research
Reflexivity	When responsibility come into consideration in the decision-making and what the objectives addressed	How the different fields integrate the dimensions of responsibility in the research activities	The university strategy toward improving responsibility in research
Responsiveness	Actions toward improving transparency beyond the activities requested by law	Actions toward improving open access of the research outputs	Assessing the capability of open access to improve the quality of research and its valorisation

The matrix can be used as a canvas of reference to represent the characteristics of the actions within the universities. This assessment shall come from an open debate aimed at understanding who put the actions toward RRI in practice, what changes the mentioned actions produced inside the research practice, and where the actions seem working in the right direction. Activity indicators can effectively become a language for RRI evaluation (Barré, 2010), to be used under a comparative perspective to allow public debated based on hard facts, through multi-actors' interactions to improve science and society relationships.

CONCLUDING REMARKS

RRI is still under development. The paper has an explorative aim to discuss how responsible research is challenging evaluation at university. Beyond the policy rhetoric, RRI concepts need to be refined both in terms of their contribution to the evidence-based policy, and their effects on the concrete life of the actors –organizations and individuals - involved. The mentioned items are at the moment mostly unexplored.

This is particularly important considering the intrinsic problems linked to the implementation of RRI in the university governance. The traditional dual orientation of the central government level and of the academics in universities in more recent time has been addressed through the introduction of new forms of hierarchical models and through rationalization processes, which are generally inspired by NPM principles. The transformation let emerge the meso-level of governance (Deans and Directors of Departments, centers, etc.) and organizational forms based on alliances between universities and groups, blurring the existing institutional boundaries. How the transformation of research universities interacts with the need to improve RRI at the governance level? Which mechanisms would lead to introduce and maintain RRI reflexivity in the knowledge production?

Another risk is the effect produced by RRI on the academic freedom and institutional autonomy when the push of research universities is toward excellence and impact. The downsize of RRI to a bureaucratic fulfilment is an actual element in existing practices in order to gain legitimacy before the society; however also the negative consequences of interpreting the openness of research projects as to the basic idea, hypothesis and methods to the public might conflict with the need to keep confidential on certain research developments.

In sum, evaluation and its formative contribution are central to manage the transition in science and society relationships that RRI implementation could realize. An approach toward democratic evaluation assumes participation of stakeholders and a number of activities to train participants, and to allow a balanced debate between them, considering also the needs of researchers and research organizations. (Patton, 2002).

However, for RRI entering the decision making process, universities need to change the governance design in order to take on board ideas, perceptions and values of different stakeholders, and to re-consider responsibility also for other functionalities such as teaching. In fact, any progress toward building a 'responsible university' must also consider how far RRI affect the education mission. Finally, how to harmonize the quest for responsibility with the autonomy (institutional and individual) in research activity is definitely an open issue.

REFERENCES

ANVUR (2015): "La valutazione della terza missione nelle università italiane. Manuale per la valutazione" available at http://www.anvur.org (27.02.2015)

Barben D., Fisher E., Selin C. and Guston D.H. (2008): "Anticipatory governance of nanotechnology: Foresight, engagement, and integration" In: Hackett EJ, Amsterdamska O, Lynch M and Wajcman J (eds) The Handbook of Science and Technology Studies. Cambridge, MA: The MIT Press, pp. 979–1000

Barré R. (2010): "Towards social robust S&T indicators: indicators as debatable devices, enabling collective learning" Research Evaluation, 19(3), 227-231

Callon, M., Lascoumes, P., Barthe, Y. (2009): Acting in an Uncertain World: An Essay on Technical Democracy. MIT Press, Cambridge, MA.

EC. (2011). DG Research Workshop on Responsible Research and Innovation in Europe.

Felt, U., Wynne, B., et al. (2007): Taking European Knowledge Seriously. DG Research. EC, Brussels

Guston, D. (2000): Between politics and science. Assuring the integrity and productivity of research. Cambridge, Cambridge University Press

Guston, D. (2012): "The Pumpkin or the Tiger? Michael Polanyi, Frederick Soddy, and Anticipating Emerging Technologies" Minerva, 50(3) 363-379

Guston D. (2014): "Understanding anticipatory governance" Social Studies of Science, 44(2) 218-242

Hessels L.K., van Lente H., Smits R. (2012): "In search of relevance: The changing contract between science and society" Science and Public Policy (2009) 36 (5): 387-401

Kuhlmann S., Edler J., Ordóñez-Matamoros G., Randles S., Walhout B., Gough C., Lindner R. (2016). Responsibility Navigator. Res-AGorA, Project, www.responsibility-navigator.eu (02052016)

Lepori B., Reale E. (2012): "S&T indicators as a tool for formative evaluation of research programs" Evaluation, vol. 18, p. 451-465

Owen R., Macnaghten P., Stilgoe J. (2012): "Responsible research and innovation: From science in society to science for society, with society" Science and Public Policy, 39, 751-756

Patton M. Q. (2002): "A Vision of Evaluation that Strengthens Democracy" Evaluation, 8(1) 125-139

Randles S. Laredo P. (2012): "Responsible Innovation and Responsible Governance (RI-RG): Reflections on Social Inquiry and Method" S.NET Conference, Twente, Netherlands 22-25 October

Reale E. (2014): "Governance" in 'The Contribution of Science and Society (FP6) and Science in Society (FP7) to a Responsible Research and Innovation. A Review', paper prepared for the EU SiS-RRI Conference, Rome, 19-20 November 2014, CNR Ed., Roma, p. 153-186

Rip A. (2003): Societal challenges for R&D evaluation, in P. Shapira and S. Kuhlman. Learning from Science and Technology Policy Evaluation, Cheltenham UK. Edward Elgar

Rip A., Joly P.B. (2012): "Emerging spaces and Governance. A position paper for EU-SPRI" Mimeo, December, Available at: http://www.eusprі-forum.eu/key_missions/exploratory_initiatives/ (06112016)

Stilgoe J., Owen R., Macnaghten P. (2013): "Developing a framework for responsible innovation" Research Policy 42, 1568– 1580

AUTHOR

EMANUELA REALE
Research Institute on Sustainable Economic Growth IRCRES - CNR
Via dei Taurini, 19 Rome, Italy 00185

Phone 0039 3339382110
e-mail: emanuela.reale@ircres.cnr.it

ESTABLISHING AN EVALUATION FRAMEWORK FOR PROMOTING GENDER EQUALITY IN R&I

SUSANNE BÜHRER, CHENG FAN, FLORIAN HOLZINGER, SYBILLE REIDL, MARTINA SCHRAUDNER, KATHARINA HOCHFELD, EVANHTIA K. SCHMIDT, JÖRG MÜLLER, DORA GROO AND RAIMUND BRÖCHLER

INTRODUCTION

EFFORTI (=Establishing an Evaluation Framework for Promoting Gender Equality in R&I) is an EU funded H2020 project that started 1st of May 2016 and will last until April 2019. The authors will present the overall aim and approach of this project.

The EFFORTI consortium consists of six partners representing a wide range of different institutional types; namely contract research organizations (Fraunhofer Society, Joanneum Research), universities (University of Catalunya, University of Aarhus), a NGO (Association of Hungarian Women in Science - NaTE) and a company (Intrasoft), distributed all over Europe (Germany, Austria, Spain, Denmark, Hungary, Luxembourg).

The aim of the project is to systematize and deepen knowledge on the scope, relevance, effectiveness and efficiency of gender equality (GE) policies on Research, Technology Development and Innovation (RTDI) by designing a common analytical framework. It will provide

- Evidence for the effectiveness of GE policies to promote gender equality and structural change in RTDI
- Evidence for the interrelations between GE in RTDI and the quality of RTDI processes and outputs.
- A smart combination of gender equality and RTDI indicators by advancing the state of the art in measuring gender equality and RTDI performances.
- A thorough contextual analysis of evaluated policies and their outputs and impacts on the national and organizational level.

EFFORTI aims to measure the progress in the area of gender equality and RTDI policy, including the stock taking and further development of tools, methods and criteria to evaluate gender equality policies in national RTDI systems. The latter will comprise the development of a robust set of guidelines that may be used at all relevant policy levels (EU, national, regional) in Europe and correspond to different policy phases (design, implementation, monitoring, evaluation, etc.). The ultimate aim of EFFORTI is to contribute to better GE policy making across Europe by analyzing a broad range of different GE policy measures with regard to their impacts on gender equality, research, innovation and competitiveness, but also on the solution of Grand Challenges and the promotion of responsible research and innovation (RRI).

THE APPROACH

In our concept and approach we combine the evaluation of gender equality policies with the most recent approaches of RTDI evaluation in order to make the best use of mutual exchange and learning. Specifically, we are going to figure out the links between policies aiming to promote gender equality - through three main objectives (more women in R&D, women in leadership positions and integration of a gender dimension in research content and curricula) - and a variety of impacts on research and innovation.

With the rise of the idea of evidence-based policy-making (e.g. Nutley et al. 2002; Solesbury 2001; Sanderson 2002), expectations have grown regarding the use of scientific evidence in policy-making. At the same time, establishing causal relationships between policy interventions and observed changes poses a theoretical challenge as well as empirical and methodological problems. One approach to address these challenges is the theory-based impact evaluation approach (TBIE): In theory-based impact evaluation (TBIE), causality is often defined as a problem of contribution, not attribution. „Why and how" questions are typically being asked instead of „how things would have been without" like counterfactual approaches do. The goal is to answer the „why it works" question by identifying the theory of change („how things should logically work to produce the desired change") behind the program and assessing its success by comparing theory with actual implementation.

The „theories" to be investigated on how gender equality and RTDI outcomes interrelate (intervention logics), which in turn link the allocation of resources to the achievement of intended results and – finally – impacts are still to be developed.[1] The actual results of GE policies will depend both on policy effectiveness and on other context variables. An essential element of policy effectiveness is the mechanisms that make the intervention work. Mechanisms are not the input-output-result chain, the logic model or statistical equations. They concern, amongst others; beliefs, desires, cognitions and other decision-making processes that influence behavioural choices and actions. Context factors are organizational structures and cultures, as well as national and regional structures, capabilities and policies. The application of a theory based impact evaluation approach will allow us to take these different levels of

1 These might be complemented by academic theories about public interventions and already existing empirical evidence from former evaluations and impact assessments.

influences on policy effectiveness - mechanisms and context - systematically into account. Furthermore it allows us to develop context sensitive and policy specific theories of change.

In conventional RTDI impacts assessments, economic impacts are the core, even if they are hardly possible to assess due to long-term effects and complex environments. But due to a shift in policy-making and programs during the last years (mission orientation, Grand Challenges etc.) it became obvious that different kinds of impact dimensions have to be considered much more, especially societal, system effects and behavioral additionality. In the EFFORTI context, social impacts seem to be particularly important, as they may include many dimensions such as new jobs, behavioral aspects like the promotion of innovation and entrepreneurial mentality, the awareness of societal needs, a better integration of the public in RTDI processes, acceptance of gender equality measures, changes in the gendered substructures of organizations or attitudes towards a better integration of gender in the innovation system.

The work is divided into different work packages: In a first step we will map the national GE and RTDI systems of the countries investigated in EFFORTI. Afterwards a first version of an evaluation framework will be developed which is validated through case studies. The final product is an evaluation toolbox.

Considering this heterogeneity between the selected countries we will conduct a thorough context analysis to describe the structure and governance of each national innovation system focusing on its relation to GE policies in RTDI and to determine how national „welfare and gender regimes" (Lewis 1999, Betzelt 2007, Pascall 2008, Plantenga 2014) structure GE in RTDI in general and the academic and scientific career opportunities of women in particular. Moreover the existing GE policies in RTDI in each country will be mapped and analysed how they are related to the overall „welfare and gender regime" and the structure of innovation systems. It will be also important to develop an understanding of evaluation cultures in each country and to map existing evaluations especially of GE policies in RTDI.

The development of a common evaluation framework for GE measures is the purpose of the next work package. The preliminary model and evaluation framework, EFFORTI 1.0., is tested and refined through case studies in the analytical validation phase. An adequate evaluation model will be validated and improved during the entire project period. The starting point is based on the purpose of the GE measures, the corresponding initiatives, their implementation activities, and their intended and unintended additional effects, i.e. objectives, inputs, throughputs, and outputs, effects and impacts. Inputs may be influenced by the policy context, throughputs by the organisational contexts and outputs, results and impacts by research team contexts. The contexts thereby influence the effectiveness of the GE measures, and how efficient GE measures are in a given national or European context.

In the logic model underlying EFFORTI 1.0., the process is seldom linear, and even though an influence of an initiative is identified, it may also include feedback, reinforcement and external political and/or system or other context influences. As a method to find systematic effect loops in the evaluations of GE initiatives, the analyses of existing GE measure evaluations and field specific guidelines will be supplemented with a comparative analysis of cases approach that uses fsQCA to identify such patterns.[2]

This process will result in a model and evaluation framework – an evaluation toolbox EFFORTI 2.0. The evaluation toolbox will most likely contain:

- a clear definition of the relevant concepts of GE policies, and evaluation subjects (e.g. RTDI and GE outputs, outcomes and impacts in the short-, mid-, and long-term);
- suggestions for suitable methods and methodologies to „measure" the different types of results;
- limitations and restrictions of methodologies and indicators;
- a definition of important contextual variables like the organizational culture and the national policy environment;
- guidelines for all phases of the design of relevant GE policies, i.e. the ex ante assessment of potential impacts, the monitoring and evaluation of outcome and impacts and on how to re-design or adopt the GE policies according to the organizational needs

The final result, the EFFORTI toolbox, allows the user to access the developed evaluation concepts and methods and learn how to apply them in their required context. The purpose of the toolbox is to provide policymakers and science managers with a broad information and consultancy tool on evaluation methods and tools for GE policies. They should find answers to questions on how to evaluate their policies in question, how to choose meaningful indicators etc. and how to proceed with the results of the evaluation, for example for the re-design of existing measures or the design of new ones. Further elements will also be accessible on the website: a toolbox with international good practices from different areas and countries.

EXPECTED RESULTS

EFFORTI will contribute to a better understanding of the impacts of current GE policies. It will help adapt GE policies and increase their efficacy, leading to an improved research intensity, productivity and responsibility and furthering the progress towards the achievement of the European Research Area. Furthermore, it will provide evidence of good practice but also concepts and tools for monitoring and evaluating GE policies and their effects on RTDI. It will therefore advance the discussion and the state of the art of measuring impacts of GE policies on RTDI by providing a comprehensive evaluation framework including an empirically tested and validated set of indicators and clear methodological guidelines on how to apply these indicators.

EFFORTI combines the theories, models and practices from GE evaluation with the most recent RTDI evaluation approaches. In particular we intend to investigate not only how GE can be improved and its effects on research and innovation outputs like number of publications and patents, **but especially RRI-related concepts like the contribution to addressing Grand Challenges, public engagement etc.**

Secondly, in order to overcome the well-known limits of conventional evaluation and impact assessment approaches, we will make use of the

2 fsQCA is especially well suited for analysis of indicators based on qualitative data and has a distinct advantage in being able to identify combinations of factors and strategies that lead to successful outcomes. Unlike linear thinking fsQCA assumes that there can be multiple configurations of strategies for reaching good outcomes and that these configurations are context dependent.

concept of theory based impact evaluation which is reflected in a sophisticated logical modelling of contributional links, the extensive consideration of the respective national and organizational framework conditions and finally a sound qualitative approach based on case studies and their validation.

In this regard, EFFORTI seeks to highlight, conceptualize and finally better understand the importance of broader systemic framework conditions for the effectiveness and efficiency of GE policies. It takes context and heterogeneity seriously. Furthermore it will provide a better understanding on how GE policies are working and achieving their impacts. Thus it will enable learning by stakeholders, policy makers and program managers.

REFERENCES

Betzelt, S. (2007), „Gender Regimes": Ein ertragreiches Konzept für die komparative Forschung, Literaturstudie, ZeS-Arbeitspapier, Bremen.

Nutley, S.; Huw D.; Walter, I. (2002): Evidence Based Policy and Practice: Cross Sector Lessons From the UK. Hrsg. ESRC UK Centre for Evidence Based Policy and Practice. Working Paper 9. London

Pascall, G. (2008), "Gender and European Welfare States", in Abrahamson, P.u. Aspalter, C. (Hg.), Understanding European Social Policy, Hong Kong.

Plantenga, J. (2014), Searching for welfare, work and gender equality, WWWforEurope Working Paper, Utrecht.

Lewis, J. (1999), Rethinking Social Policy: Gender and Welfare Regimes, IWM Working Paper, Wien.

Sanderson, I. (2002): Evaluation, Policy Learning, and Evidence Based Policy Making, Public Administration 80(1), 1-22

Solesbury, W. (2001): Evidence Based Policy: Whence it Came and Where it's Going. Hrsg. ESRC UK Centre for Evidence Based Policy and Practice. Working Paper 1. London

AUTHORS

SUSANNE BÜHRER
Fraunhofer ISI
Breslauer Str. 48, 76139 Karlsruhe
Germany
Tel.: +49-721-6809-148

CHENG FAN
Fraunhofer ISI
Breslauer Str. 48, 76139 Karlsruhe
Germany

FLORIAN HOLZINGER
Joanneum Research
Sensengasse 1, 1090 Vienna
Austria

SYBILLE REIDL
Joanneum Research
Sensengasse 1, 1090 Vienna
Austria

MARTINA SCHRAUDNER
Fraunhofer CeRRI
Hardenbergstr. 20, 10623 Berlin
Germany

KATHARINA HOCHFELD
Fraunhofer CeRRI
Hardenbergstr. 20, 10623 Berlin
Germany

EVANHTIA K. SCHMIDT
Aarhus University
Bartholins Allé 7, 8000 Aarhus C
Denmark

JÖRG MÜLLER
Universitat Oberta de Catalunya
Av. Carl Friedrich Gauss 5, 08860 Castelldefels
Spain

DORA GROO
Association of Hungarian Women in Science
Napraforgo u 17, 1021 Budapest
Hungary

RAIMUND BRÖCHLER
INTRASOFT Intl. S.A.,
2B Rue Nicolas Bové, 1253 Luxembourg
Luxembourg

IMPACT ASSESSMENT AND GRAND CHALLENGES

MAGNUS GULBRANDSEN, JAKOB EDLER AND JORDI MOLAS-GALLART

The 2016 Open Evaluation conference aims to provide knowledge foundations for policy evaluations within research, technology and innovation policy, and there is a special Theme on societal impact and societal challenges (Theme number 2) which this paper intends to contribute to. Our paper fits in between the societal challenges sub-theme and the impacts of agricultural research sub-theme. We take a somewhat broader outlook on impact than just related to agricultural research. At the same time we acknowledge that agricultural research is a particularly interesting area due to its long traditions for impact assessment and the promising recent methodologies for impact assessment developed especially related to the French INRA institute (Joly et al. 2015; Gaunand et al. 2015). It can also easily be tied to societal challenges related to health, food supply and more.

A long-standing debate in science policy has been how to prioritise between fields of science and between research institutions (Weinberg 1962a and b). In a time of increasing research budgets such as the first two-three decades after WW2, this prioritisation was relatively easy, leading to a fairly stable and path-dependent balance between long-term research steered by researchers' own agendas and research steered by various societal needs. Evaluations in this context largely served to monitor the system and identify fields and organisations where extraordinary policy efforts could improve quality or reward high performance.

But since the first post-war crisis in the 1970s and not least in the current one, the emphasis on hard priorities has become stronger. The dominating current rationale is that of "grand challenges" denoting fundamental international and shared problems where research and innovation are seen as major activities in large-scale concerted societal efforts (Kuhlmann & Rip 2015). Research evaluations have become a clearer instrument in distribution of resources between fields and organisation, and they have increasingly included measures of the effect and impact of research in society rather than intra-scientific characteristics only. Impact is high on the agenda and has become a central legitimisation for research and innovation support. But grand or societal challenges are often also very broad and fuzzy, sometimes barely operationalised beyond terms such as "aging population", "global health" and "climate change". There is a need to disentangle this further in order to make sensible evaluation designs.

RESEARCH AIM AND QUESTIONS

This paper aims to instigate a comprehensive and conceptual discussion of the relationship between impact (assessment) and grand challenges. We want to put forward a number of propositions for debate in the research evaluation community, and we want to shape a future-oriented research agenda about impact. Our context is a joint project with long-term (8 year) funding set up to conduct new and ambitious empirical research on research impact. The main ideas are to bridge the gap between diverse impact assessment frameworks and to carry out longitudinal studies highlighting in particular the absorptive capacity and uptake of research by users.

Even though there is plentiful knowledge about diffusion, use and impact of research and its results, there is still a widespread and well-founded belief that analysts have not cracked the problem of research impact assessment in a way that it is useful for policy makers. We argue that the main reason for this is fragmented research on the nature and modes of impact (different communities/perspectives, ad-hoc evaluations) and that fragmentation has restricted knowledge accumulation. Through discussions in the RTI evaluation community, drawing on examples and experiences from many different countries, we will strive to find new ways of reducing the fragmentation in the field of impact studies and new ways of tying impact to the framework of grand challenges.

DEFINITIONS

Research impact and research impact assessment is a broad area of research which over the last fifty years has generated substantial knowledge about different kinds of impact and the different ways investments in research influence economic, political, social, cultural and environmental developments (Godin & Dore 2004, Donovan 2011). Since investment in public and private research can have, in theory, many different impacts in very different sectors of society, the conceptual and methodological approaches to research impact assessment are correspondingly heterogeneous. The current state-of-the-art is to address impact through a combination of quantitative and qualitative data and to see impact as a heterogeneous phenomenon related to multiple stakeholders (Bornmann 2013).

THEORETICAL FRAMEWORKS

Roughly speaking, four different communities have been engaged in impact studies. The first, oriented at economics of R&D, has looked mainly at economic effects and has attempted to study the returns on public and private investment in research and development (R&D). The impact that private sector R&D has on the economic performance of firms and the economy at large has been a persistent theme both in public debate and in research on the economics of R&D and innovation (Mansfield 1990, Salter & Martin 2001). Empirical research indicates that social returns from R&D investments tend to be substantially higher than the potentials for the firm (Griliches 1995, Jones & Williams 1998). This is related to the partly non-appropriable and public nature of technological knowledge that leads to spillover effects on later research and by increasing productivity of other economic activities (Griliches, 1995; Geroski, 1995; Jones and Williams, 1998; Hall et al. 2010).

We define social impact studies and research evaluation as the second relevant community. Its starting point is the understanding that research impact is heterogeneous and denotes more than "economic benefits" and involves many different stakeholders (Bornmann 2013). In addition to impact heterogeneity, other central challenges are related to latency, causality and attribution (Buxton 2011; Martin 2007). A range of new approaches to research impact assessment have been designed and implemented to deal with such challenges. The "Payback Framework" looks at different types of impact over time (Buxton & Hanney 1996, Donovan & Hanney 2011), the SIAMPI effort focuses on "productive interactions" between researchers and external stakeholders (Spaapen & van Drooge 2011; Molas-Gallart & Tang 2011), and the ASIRPA approach combines qualitative and quantitative data to gain insights into how different forms of impacts often appear together for different beneficiaries (Joly et al. 2015; Gaunand et al. 2015).

A third relevant community has studied knowledge exchange and science-based innovation. Here, a central perspective is the diversity of channels that academics and stakeholders use to maintain interactions and communicate with each other (Perkmann et al. 2013; Abreu & Grinevich 2013; Thune et al. 2015; Olmos-Peñuela et al., 2014). Empirical investigations indicate that the volume of knowledge exchange activities is relatively similar across different fields of science (Hughes & Kitson 2012). However, the channels or tools used for knowledge exchange differ markedly by fields of science, as do the types of stakeholders who are perceived as the most important partners (Abreu & Grinevich 2013; Thune et al. 2015; Olmos-Peñuela et al., 2014; Ramos-Vielba et al., 2015; Upton et al. 2014).

Finally, there is a diverse community that has conceptualised how specific products and technologies emerge over a long time period (e.g. Dosi 1982; Blume 1992; Bijker 1995). Technical change is seen as a socio-technical transformative process that involves not only changes in technology and the related scientific knowledge base, but also transformation in the social context in which the technology is embedded. It has for example been shown how scientific breakthroughs in medicine influence and are influenced by learning in medical practice and by new technologies and products, conceptualised as three "co-evolving pathways" (Morlacchi & Nelson 2011).

METHODOLOGICAL APPROACH AND OUTCOMES

With these four communities in mind, we aim to set up a number of tensions and questions for debate concerning impact assessment and grand challenges. A non-exhaustive preliminary list is the following:

- There seems to be a major gap between qualitative and quantitative approaches to impact assessment; is it possible to find some new approaches to combining the two? The ASIRPA approach seems promising but is mainly tested on the impact of one specific research organisation in one specific area (agriculture), and scaling the methodology up may pose problems. Large-scale databases and big data approaches may provide a link.
- What is the relationship between impact and grand challenges; for example, can high impact with respect to some of the grand challenges (particular health issues, environmental issues and so on) be traced to research that was funded specifically to deal with these challenges, or to other types and areas of research? How can non-intended (positive and negative) impacts of grand challenges-legitimised research be understood and dealt with?
- Responsible research and innovation (RRI) is an approach that seems to be favoured in policy communities discussing grand challenges; the approach entails discussing effects and impacts between many different stakeholders in an early stage of the research and innovation process. Can impact studies learn from the RRI approach and vice versa?

These are examples of issues we want to highlight and discuss in the final paper and at the Open Evaluation conference – and these are issues that have both a fundamental interest for the researchers who study science, technology and innovation and for the policymakers and other practitioners who work in this area.

REFERENCES

Abreu, M. & Grinevich, V. 2013. The nature of academic entrepreneurship in the UK: widening the focus on entrepreneurial activities, Research Policy, 42:408-422

Bijker, W.E. 1995. Of Bicycles, Bakelites, and Bulbs. Toward a Theory of Sociotechnical Change. Boston, MA: MIT Press.

Blume, S. S. 1992. Insight and industry: on the dynamics of technological change in medicine. Boston, MA: MIT Press.

Bornmann, L. 2013. What is societal impact of research and how can it be assessed? A literature survey, Journal of the American Society for Information Science and Technology, 64(2):217-233.

Buxton, M. & Hanney, S. 1996. How can payback from health services research be assessed? Journal of Health Services Research & Policy, 1(1):35-43.

Donovan, C. 2011. State of the art in assessing research impact: introduction to a special issue. Research Evaluation, 20(3):175-179.

Donovan, C., & Hanney, S. 2011. The 'Payback Framework' explained, Research Evaluation, 20(3):181-183.

Douthwaite, B., Kuby, T., van de Fliert, E., & Schulz, S. 2003. Impact pathway evaluation: an approach for achieving and attributing impact in complex systems. Agricultural systems, 78(2):243-265.

Gaunand, A., Hocdé, A., Lemarié, S., Matt, M. & de Turckheim, E. 2015. How does public agricultural research impact society? A characterization of various patterns, Research Policy, 44:849-861.

Geroski, P. 1995. "Markets for technology. Knowledge, innovation and appropriability." In Stoneman, P. (Ed.): Handbook of the Economics of Innovation and Technological Change. Oxford: Basil Blackwell.

Godin, B. & Doré, C. 2004. Measuring the impacts of science: beyond the economic dimension. History and sociology of R&D statistics working paper, not numbered.

Griliches, Z. 1991. The search for R&D spillovers. NBER Working Paper 3768.

Griliches, Z. 1995 "R&D and productivity: econometric results and measurement issues." In Stoneman, P. (Ed.), Handbook of the Economics of Innovation and Technical Change. Oxford: Basil Blackwell.

Gulbrandsen, M., D. Mowery & M. Feldman, 2011, Heterogeneity and university-industry relations: introduction to the special section, Research Policy, 40:1-5.

Hall, B.H., Mairesse, J. & Mohnen, P. 2010. Measuring the returns to R&D. Handbook of the Economics of Innovation 2, pp. 1033-1082.

Hughes, A. & Kitson, M. 2012: Pathways to impact and the strategic role of universities: new evidence on the breadth and depth of university knowledge exchange and the factors constraining its development. Cambridge Journal of Economics, 36:723-750.

Joly, P. B., Gaunand, A., Colinet, L., Larédo, P., Lemarié, S., & Matt, M. 2015. ASIRPA: a comprehensive theory-based approach to assessing the societal impacts of a research organization. Research Evaluation, 24:1-14.

Jones, C.I. & Williams, J.C. 1998. Measuring the social return to R&D. The Quarterly Journal of Economics, 113(4):1119-1135.

Mansfield, E. (1991). Academic research and industrial innovation. Research policy, 20(1):1-12.

Martin, B.R. 2007. Assessing the impact of basic research on society and the economy. Paper presented at the Rethinking the impact of basic research Conference, Vienna, 11 May.

Molas-Gallart, J. & Tang, P. 2011. Tracing 'productive interactions' to identify social impacts: an example from the social sciences. Research Evaluation, 20(3):219-226.

Morlacchi, P., & Nelson, R. R. 2011. How medical practice evolves: Learning to treat failing hearts with an implantable device, Research Policy, 40(4):511-525.

Morton, S. 2015. Progressing research impact assessment: A contributions approach. Research Evaluation, august 2015:1-15.

Olmos-Peñuela, J., Molas-Gallart, J. & Castro-Martinez, E. 2014. Informal collaborations between social sciences and humanities researchers and non-academic partners, Science and Public Policy, 41(4):493-506.

Perkmann, M. et al. 2013. Academic engagement and commercialisation: A review of the literature on university-industry relations, Research Policy, 42(2):423-442.

Salter, A.J. & Martin, B.R. 2001. The economic benefits from publicly funded R&D: a critical review, Research Policy, 30(3):509-532.

Spaapen, J. & van Drooge, L. 2011. Introducing 'productive interactions' in social impact assessment, Research Evaluation, 20(3):211-218.

Thune, T. , Reymert, I. & Gulbrandsen, M. 2015. Universities and external engagement activities: Particular profiles for particular universities? Forthcoming in Science and Public Policy.

Upton, S., Vallance, P. & Goddard, J. 2014. From outcomes to process: evidence for a new approach to research impact assessment, Research evaluation, 23:352-365.

AUTHORS

MAGNUS GULBRANDSEN
Corresponding author
Professor
TIK Centre for Technology, Innovation and Culture, University of Oslo
PO Box 1108 Blindern
NO-0317 Oslo
Norway
Email: magnus.gulbrandsen@tik.uio.no
Phone: (+47) 90 72 59 78

JAKOB EDLER
TIK Centre for Technology, Innovation and Culture, University of Oslo
PO Box 1108 Blindern
NO-0317 Oslo
Norway

JORDI MOLAS-GALLART
TIK Centre for Technology, Innovation and Culture, University of Oslo
PO Box 1108 Blindern
NO-0317 Oslo
Norway

WIFAS - A MODEL AND METHODOLOGY TO CAPTURE THE SOCIAL EFFECTS CAUSED BY SUBSIDY PROGRAMMES TO ADVANCE MOBILITY RESEARCH

PETER KAUFMANN, LAURENZ WOLF, ALEX SCHUBERT AND ALEXANDER NEUMANN

We can observe an increasing role for estimating impacts of RDTI subsidy programmes over the years. Because of the rising need for legitimacy over the use of public money, policy makers have been asking for more impact assessments as a basis for evidence-based policy making. This has also resulted in a new approach to handling public finances, which is subsumed under the heading of „results-oriented management of public finances", and we have also seen a rise of the demand to respond to pressing societal challenges. This taken together means that more detailed information of the whole spectrum of impacts needs to be provided.

Thus, evaluations of RDTI policies have increasingly covered impacts on innovation and competitiveness, or cooperation structures between stakeholders like industry-science links. The results of these evaluations have also played a pivotal role in discussing the further direction of science policy. More recently, we can observe an increasing pressure to include also further dimensions like impacts on the environment, some of which are now increasingly incorporated. The least developed dimension of the 'triple bottom line' is the social impact dimension of research into RDTI, of which some are mentioned under the themes 'passenger mobility' and 'freight mobility' in the portfolio of the Federal Ministry for Transport, Innovation and Technology (bmvit). These are direct effects like the accessibility of the transport system or ensuring the provision of goods and services, and more indirect effects like social cohesion or health impacts.

In the light of the above, this study aimed to develop a conceptually and empirically sound intervention logic to capture the potential social impacts caused by subsidy programmes to advance mobility research. By doing so, we aimed to answer the following questions: (a) which relevant social effects and/or impacts are caused by project results of mobility research programmes, and (b) which methods and indicators can be used to represent those social effects. The draft model was fed back to the research community via expert interviews, and a further developed version empirically tested using case studies for selected research projects funded under the past and current mobility research programmes.

Starting from the state of the art of capturing social effects in different research fields, the study not only developed the conceptual model for the specific policy purpose, but also an appropriate methodological tool to capture social impacts in the sense of a "qualitative scoping" (EC 2015), and made some initial allocation of potential indicators to the individual impact dimensions based on the conceptual model. The model is based on an initial idea by Jones and Lucas (2012) to separate social effects from distributional effects and conceptualise the latter to be a cross cutting dimension through all other social effects (and indeed also economic and environmental). On that basis, we tried to enhance the analytical concept and applied it to the field of personal and goods mobility research/policy.

Further, the methodological tool was successfully tested and can be applied in the future in a wider context. Defining a coherent indicator framework is more challenging, because the whole breath of the diverse projects funded needs to be covered to decide whether specific indicators can be generally applied at the project level, and thus aggregated to the thematic and programme levels. This needs to be the topic of a future study.

Basically, we follow the approach suggested by the European Commission (EC 2009, 2015) to define a relatively simple process with three steps: (1) Identification of impacts; (2) Qualitative assessment of the more significant impacts; (3) In-depth qualitative and quantitative analysis of the most significant impacts. In this project, we developed a theoretically sound basis for a 'qualitative scoping' to respond to step one and two. The results from the projects can be aggregated to thematic fields and the whole subsidy programme. This result can form the basis for a decision to concentrate in a further step to investigate the main and/or most interesting social impacts in more detail.

Partly following the argumentation of an expert commission for the EC and OECD (2015) that „adopting a measuring process rather than imposing specific metrics or indicators" is a fruitful way forward, we also see that public authorities do need measurable indication of which effects their policies are initiating. Still, one needs to be cautious about a possible over-quantification of potentially spurious effects. This is why we suggest using quantitative indicators only for well-established links of causation, where attribution is credible and these are preferably accompanied by qualitative assessments to support the argument. Otherwise, qualitative indication of effects seems to be the more promising path to follow.

In summary, the empirically validated model does not only contribute to the conceptually underdeveloped question of which and how to capture social effects of RDTI mobility programmes, but can also be adapted to other thematic fields without too much effort.

EXCERPT OF THE MOST RELEVANT REFERENCES

(altogether around 200)

Adelle, C., Weiland, S. (2012). Policy assessment: the state of the art. Impact Assessment and Project Appraisal, 30(1): 25-33

Atkins (2010). Assessing Social and Distributional Impacts in Transport Scheme Appraisal and Evaluation. Pilot Phase: Final Report by Atkins Transport Planning & Management for the Department for Transport (DfT), London

Delbosc, A., Currie, G. (2011). Transport problems that matter - social and psychological links to transport disadvantage. Journal of Transport Geography, 19(1): 170-178

European Commission (2009). Impact Assessment Guidelines. SEC(2009) 92

European Commission (2015). Better Regulation Guidelines, Commission Staff Working Paper, SWD(2015) 111 final

European Commission/OECD (2015). Policy brief on social impact measurement for social enterprises, Luxembourg

Forkenbrock, D.J., Weisbrod, G.E. (2001). Guidebook for Assessing the Social and Economic Effects of Transportation Projects, NCHRP Report 456 for the Transportation Research Board, Washington

Geurs, K. T, Boon, Wouter and Van Wee, Bert (2009). 'Social Impacts of Transport: Literature Review and the State of the Practice of Transport Appraisal in the Netherlands and the United Kingdom' Transport Reviews, 29(1): 69–90

Hills, D. Junge, K. (2010). Guidance for Transport Impact Evaluations - Choosing an evaluation approach to achieve better attribution. The Tavistock Institute in consultation with AECOM for the Department for Transport (DfT), London

IAIA (2015). Social Impact Assessment: Guidance for assessing and managing the social impacts of projects. International Association of Impact Assessment

Jones P. Lucas, K. (2012). The social consequences of transport decision-making: clarifying concepts, synthesising knowledge and assessing implications. Journal of Transport Geography 21, 4-16

Litman, T. (2014). Well Measured - Developing Indicators for Sustainable and Livable Transport Planning, Victoria Transport Policy Institute

Litman, T. (2015). Evaluating Public Transit Benefits and Costs, Best Practice Handbook, Victoria Transport Policy Institute

Lucas, K. (2012). Transport and Social Exclusion: Where Are We Now? Transport Policy 20, 105-113

Maas, K., Liket, K. (2011). Social Impact Measurement: Classification of Methods. Environmental Management Accounting and Supply Chain Management, 171-202

Markovich, J., Lucas, K. (2011). The Social and Distributional Impacts of Transport: A Literature Review, Working Paper No. 1055, Transport Studies Unit, School of Geography and the Environment, University of Oxford

OECD (2010). Guidance on Sustainability Impact Assessment, Paris

Parkhurst, G., Shergold, I. (2009). Literature Review - The Treatment of Social and Distributional Impacts in Appraisal Evaluation - Final Report. University of the West of England

Simsa, R. Rauscher, O. Schober, C. Moder, C. (2014). Methodological Guideline for Impact Assessment, FP 7 project IMPACT, working paper No. 01/2014

Slootweg, R., Vanclay, F. and van Schooten M. (2001). Function evaluation as a framework for the integra-tion of social and environmental impact assessment. Impact Assessment Project Appraisal 19(1): 19-28

Social Exclusion Unit (2003). Making the Connections: Final Report on Transport and Social Exclusion. London: ODPM

Stanley, JK. Hensher, DA., Stanley, JR., Vella-Brodrick, D. (2011). Mobility, social exclusion and well-being: Exploring the links. Transportation Research Part A 45: 789-801

Tetlow, M. Hanusch, M. (2012). Strategic environmental assessment: the state of the art. Impact Assessment and Project Appraisal 30(1): 15-24

Titheridge, H., Christie, N., Mackett, R., Oviedo Hernández, D., Ye, R. (2014). Transport and Poverty: A review of the evidence, technical report, University College London: 1-47

Turnpenny, J., Radaelli, CM., Jordan, A., Jacob, K. (2009). The policy and politics of policy appraisal: emerging trends and new directions. Journal of European Public Policy, 16(4): 640-653

Vanclay F. (2003). International Principles For Social Impact Assessment, Impact Assessment and Project Appraisal, 21(1): 5-11

Vanclay, F. (2002). Conceptualising social impacts. Environmental Impact Assessment Review 22: 183-211

World Bank (2006). Social Analysis in Transport Projects: Guidelines for Incorporating Social Dimensions into Bank-Supported Projects. Social Analysis Sector Guidance Note Series, Social Development Department of the World Bank

World Bank (2007). Tools for Institutional, Political, and Social Analysis of Policy Reform. A Sourcebook for Development Practitioners.

AUTHORS

PETER KAUFMANN
LAURENZ WOLF
ALEX SCHUBERT
ALEXANDER NEUMANN

KMU Forschung Austria / Austrian Institute for SME Research
Gußhausstraße 8, 1040 Vienna
Austria

Netwiss OG
Hohe Warte 46, 1190 Vienna
Austria

JOINT EVALUATION FOR JOINT GOVERNANCE OF CHALLENGE-ORIENTED RESEARCH

LEONIE VAN DROOGE AND JASPER DEUTEN

SUMMARY

We propose a novel way to understand evaluation of challenge-oriented research. We argue that challenge-oriented research requires specific modes of governance and research evaluation. Challenge-oriented research goes beyond conventional thematic or mission-oriented research programming. It concerns broader transformations in society, for instance the transitions towards sustainable energy systems or sustainable agriculture. It necessarily involves a broad range of stakeholders who may have different views on what the problem is and how it should be approached. Challenge-oriented research is part of a complex, non-linear, long-term, open-ended and contested transformation journey. This should be reflected in the governance and evaluation of this mode of research. We argue that 'joint' and 'in itinere' evaluation is crucial to learn how and what research contributes to broader, systemic transformations in society. It is also a necessary ingredient in building trust between various research funders, multiple research performers, users and other stakeholders that research helps society to move forward a step in the (open-ended) transformation journey.

For this type of research, traditional ways of research evaluation do not suffice. New evaluation methods and practices have been developed over the last years. But so far, the experience with these methods for the evaluation of challenge-oriented research is limited. We argue that understanding evaluation as a joint governance process is key. We present a number of projects concerning evaluation of challenge-oriented research and the lessons learnt.

INTRODUCTION

Science is increasingly called upon to contribute to finding and developing solutions for grand societal challenges. Scientific research is perceived as an integral part of innovation journeys (Van de Ven et al, 2008) towards innovative solutions. It has become more important for scientists to demonstrate how they help shape such innovation journeys towards finding solutions for societal challenges and problems.

Typically, societal challenges require broader transformations in society, for instance transitions towards sustainable energy systems or sustainable agriculture. Individual research projects and programmes need to contribute to such broader transitions. Moreover, researchers need to show the quality and relevance of their research to a broad range of stakeholders who may have different views on what the problem is and how it should be approached.

New modes of scientific research have emerged that allow for a challenge-oriented approach and a more effective incorporation and engagement of society (Gibbons et al. 1994, Nowotny et al. 2001). Science is becoming more 'open' for the involvement of other academic disciplines, research institutes, businesses and societal organisations. Challenge-oriented research is typically transdisciplinary and based upon collaboration and co-creation by various actors. This differs from mission oriented research, such as the Man on the Moon or the Manhattan projects. It refers to wicked problems, such as poverty in the ghetto. According to Nelson (1977), putting a man on the moon was a 'simple' problem: there was general agreement on the problem, there was a single owner of the problem, and there was a single technological solution. The ghetto is a wicked problem: every stakeholder has its own visions and perceptions of the problem, there is no clear owner of the problem, and there are no clear solutions nor can solutions easily be translated from one ghetto to the next. Challenge-oriented research is part of a complex, non-linear, long-term, open-ended and contested transformation process.

As part of these new modes of research, new governance arrangements have been developed to guide and steer research towards societal needs. For example when a broader and more diverse range of stakeholders is involved in agenda and priority setting. The Dutch National Research Agenda (2016) is an interesting example, where citizens were asked to contribute research questions as well. Medical charities in the Netherlands have developed practices to include end-users (patients, doctors, carers) in agenda setting and project selection. Funding and spending arrangements increasingly include a more heterogeneous mix of parties. This includes co-funding by various public and private sector actors as well as allocation to multi-party (public-private) consortia with complementary research and innovation actors. Finally, new arrangements for knowledge sharing and intellectual property have been developed (open access, open data), that still have to find a balance between protection and sharing.

New governance arrangements for challenge-oriented scientific research require new methods to evaluate the quality and impact of research. Conventional evaluation methods that primarily focus on scientific excellence and that are based on peer-review and scientometrics (e.g. journal impact factors, citation impact analysis), do not suffice. Scientific excellence is in most, if not all, cases only one element, or one type of activity, or one quality of the research.

CHALLENGE ORIENTED RESEARCH REQUIRES A NEW EVALUATION APPROACH

Evaluation of challenge-oriented research relates to both the knowledge produced and its subsequent use, as well as the process of knowledge co-creation. It takes into account how, and to what extent, research contributes to shaping promising and/or effective innovation journeys that further societal transformations. In other words, challenge-oriented research requires a sophisticated theory of change.

Over the years research evaluation methods and systems have been designed and implemented that include stakeholders and that acknowledge the contribution of research to societal challenges (Donovan 2011, Bornmann 2013). Examples in practice vary and refer to ex ante as well as ex post evaluations. From ex ante evaluation of broader impacts in the US (National Science Foundation (NSF) proposals), pathways to impact in the UK (Research Councils UK (RCUK) proposals) and knowledge utilization in the Netherlands (research council NWO proposals), to ex post evaluations of social impact in the UK (Research Excellence Framework (REF) 2014) and relevance to society in the Netherlands (Standard Evaluation Protocol (SEP) 2015-2021).

In general, broader impact of research has become one of the key evaluation criteria in research evaluation by research funding agencies. In most of these cases the research is not driven by a specific challenge. It is up to the researcher to formulate a challenge or impact. At best, the funder has indicated the types of impacts, processes or stakeholders that are within the limits of what the funder has defined as impact. This is in line with the key condition of academic freedom. For challenge oriented research, however, the stakeholders involved are engaged in a joint innovation journey. The impact they aim for, is jointly decided and relates to the challenge. A theory of change relating to this impact can play a central role in evaluation.

THE CLOSE RELATION BETWEEN EVALUATION AND GOVERNANCE: THEORY OF CHANGE

At present, the question of what impact to evaluate, is often answered by proposing indicators that are available and quantifiable, such as patents or spin-offs. The challenge is to ensure that a realistic perspective on impact forms the basis of an evaluation. This is a joint effort, that includes the challenge that drives the funder(s). A related challenge is the development of adequate evaluation criteria, evaluation questions, indicators and methods. Again, this is a joint effort. A final challenge is to organise the involvement and engagement of multiple stakeholders in evaluation. This goes well beyond extended peer review. The consequence is that one has to take into account the variety of interests, possibly conflicting, of all included.

So the picture becomes even bigger. From evaluation of research excellence alone – that can be done through peer review – through evaluation of societal impacts – with the inclusion of stakeholders – to evaluation as a joint process with all involved – including the funder – and with a central focus on the challenge. Evaluation thus relates to far more than research alone.

Looking at evaluation this way, evaluation and governance are very similar. That might sound uninviting to some, but Hill and Lynn (2005) propose a non-hierarchical form of network-governance when they state that "governance as an organizing concept for public management reform reflects a widespread, though not universal, belief that the focus of administrative practice is shifting from hierarchical government toward greater reliance on horizontal, hybridized, and associational forms of governance". Kuhlmann and Rip (2014) call for a tentative concept of governance for challenge oriented research, that is preliminary (temporarily limited). They argue that learning processes are key. A theory of change can support this form of governance.

A theory of change (Rogers, 2014) explains how an impact is understood to come about. It is a shared narrative concerning the causal relation between inputs, activities, outputs and impact. It is a joint understanding of the innovation journey and is best developed by all involved and affected. This includes research funders, research performers, intermediate and end users of research results. Note that research funders have a stake regarding the challenge that they have addressed in their funding instrument. They can be regarded as spokesperson for the challenge addressed.

Developing and using a theory of change can be regarded as a horizontal form of governance. From the theory of change, a number of evaluation or monitoring questions can be identified (Spaapen and Van Drooge, in preparation). These can relate to elements of the theory of change for which there is no evidence yet (Rogers, 2014). Activities or outputs, when understood as part of the theory of change or of the innovation journey, can serve as indicators in itinere, or on the way. They can be used to monitor the progress and route in the course of the project. In case the results differ from what was expected the theory of change can support learning. As a consequence of the learning, the theory of change can be adjusted. Douthwaite (2016) provides an example of an adjustments of a theory of change.

EXAMPLES OF EVALUATION AND GOVERNANCE

One approach that seems particularly suited for the evaluation of challenge-oriented research is Participatory Impact Pathways Analysis (PIPA). Theory of change is the central element in PIPA. It was developed from earlier ideas in programme theory and pioneered within the Consultative Group on International Agricultural Research (CGIAR) Challenge Programme on Water and Food (Douthwaite et al. 2007a; Douthwaite et al. 2007b from WorldFish). PIPA has been applied in a number of different contexts, originally to plan and monitor the impact of research for development projects, but also for other types of challenge-oriented research. Not all experiences with PIPA were entirely satisfactory, see for instance Spaapen and Van Drooge (2015) and Triomphe et al, (2015). We have used the lessons from these experiences in two projects that we have been involved in.

We developed an evaluation protocol for the monitoring and evaluation of a number of applied research organisations in the Netherlands.

The development of the protocol involved extensive discussions with the ministries involved (funders) as well as with the research organisations. In the previous protocol, the ministries were at a great distance of the evaluation. In this protocol, the ministries are still not directly involved in the evaluation itself, but the process is designed in such a way that their goals and challenges play a crucial role during evaluation (Deuten et al, 2015).

Medical charities in the Netherlands feel a growing pressure from sponsors and patients to show how their activities have an impact. They have developed practices to improve the focus in research projects and programs on societal impact, especially in the phase of agenda setting and project selection. However, they still experience that once a project is approved, many of the researchers are more focussed on academic excellence than on societal impact. Some charities have indicated the need to redefine their role in the phase of research. We have developed a joint workshop for research managers of the charities and researchers in order to develop new forms of governance, specifically new evaluation and monitoring approaches, for the challenge oriented research projects of the charities.

A final experience that we will reflect upon is an effort that we haven't been involved in directly. It illustrates how governance and evaluation go hand in hand. A major mission oriented research organisation in the Netherlands has changed the way research and researchers are evaluated. (Benedictus and Miedema, 2016) The mission of the organisation is central to the evaluation. Research excellence is still an important criterion, but contrary to before, it is by far not the only criterion. The development of this new approach included discussions and inclusion of researchers, as well as major stakeholders, users as well as funders.

REFERENCES

Benedictus, Rinze and Frank Miedema (2016) Redefine excellence: Fix incentives to fix science. Nature Volume 538, 435-455

Bornmann, L. (2013): "What is societal impact of research and how can it be assessed? a literature survey" Journal of the American Society for Information Science and Technology Volume 64 (2), 217–233

Deuten, Jasper, Lionne Koens, Patricia Faasse and Barend van der Meulen (2015): Protocol voor de monitoring en evaluatie van de Toegepast Onderzoeksorganisaties in Nederland Den Haag: Rathenau Instituut

Donovan, Claire (2011): "State of the art in assessing research impact: introduction to a special issue" Research Evaluation 20 (3): 175-179

Douthwaite, B. et al (2007a): 'Participatory Impact Pathways Analysis: A Practical Application of Program Theory in Research-for- Development', Canadian Journal of Program Evaluation 22 (2) 127-159

Douthwaite, B., Schulz, S., Adetunji, S. O., and Ellis-Jones, J. (2007b): "Impact pathway evaluation of an integrated Striga hermonthica control project in Northern Nigeria" Agricultural Systems 92: 201-222

Douthwaite, B. (2016) THEORY OF CHANGE | Using theory of change and impact pathways to leverage results. https://www.youtube.com/watch?v=YcyhrlZxo7Y, assessed 19 October 2016

Drooge, L. van, Jong, S. de et al (2013) Twenty years of research evaluation (Facts & Figures 9) Den Haag: Rathenau Instituut

Dutch National Research Agenda (2016)
http://www.wetenschapsagenda.nl/national-science-agenda/?lang=en

Gibbons, m., C. Limoges, H. Nowotny, S. Schwartzman, P. Scott and M. Trow (1994): The New Production of Knowledge: The Dynamics of Science and Research in Contemporary Societies. London, Sage

Hill, Carolyn J., and Laurence E. Hill (2005): "Is Hierarchical Governance in Decline? Evidence from Empirical Research" Journal of Public Administration Research and Theory Vol. 15, no. 2

Edwin Horlings, Stans van Egmond, Rosalie Belder, Leonie van Drooge (2015) Ontwikkeling van een visie op de rol van de Samenwerkende GezondheidsFondsen in het wetenschaps- en innovatiebeleid. Den Haag: Rathenau Instituut.

Kuhlmann, S. & Rip, A. (2015): The challenge of addressing Grand Challenges. In: von Schomberg, R. (ed.): The Future of Research and Innovation. European Commission (in press)

Nelson, R. R. (1977). The moon and the ghetto. New York: Norton

Nowotny, Helga, Peter Scott and Michael T. Gibbons (2001): Re-Thinking Science: Knowledge and the Public in an Age of Uncertainty. London, Wiley

Rogers, P. (2014). Theory of Change, Methodological Briefs: Impact Evaluation 2. UNICEF Office of Research, Florence.

Spaapen, J. B. and L. van Drooge (2015): "Towards a network approach of research evaluation" in INRA: Impacts of agricultural research - an approach of societal values. Paris, Nov 2015

Triomphe B. (2015): "Towards a generic, comprehensive and participatory approach for assessing the impact of agricultural research in developing countries" at INRA: : Impacts of agricultural research - an approach of societal values. Paris, Nov 2015

Van de Ven, Andrew, Douglas Polley, Raghu Garud, Sankaran Venkataraman (2008): The Innovation Journey Oxford: Oxford University Press

AUTHORS

LEONIE VAN DROOGE
Corresponding author
Rathenau Institute
Anna van Saksenlaan 51, 2593 HW The Hague
the Netherlands
e-mail: l.vandrooge@rathenau.nl
+31648199823

JASPER DEUTEN
Rathenau Institute, Anna van Saksenlaan 51, 2593 HW The Hague
the Netherlands

ASSESSING AND EVALUATING NEW MISSION-ORIENTATED R&D PROGRAMS: REQUIREMENTS, FRAMEWORKS AND A REVIEW OF RECENT EXPERIENCES

K. MATTHIAS WEBER AND WOLFGANG POLT

INTRODUCTION

In recent years, and in an increasing number of countries we could observe a shift from structure-oriented R&D funding programs that were dominant in the past 20 years, aiming to enhance the ability of systems to generate innovation "per se", to funding initiatives that are supposed to help tackle major, often long-term, societal challenges. While these "new" mission-oriented R&D programs (MOPs) may share an explicit thematic goal-orientation with the "old" mission-oriented programs of sixties, they are not solely guided by technological, but predominantly by societal targets.

With this change of the purpose of R&D, the requirements for their evaluation have equally changed. To this ads, triggered not least by the European Court of Auditors, a stronger emphasis put by public authorities on the ex-ante assessment of expected or likely impacts of policy initiatives, which complements the well-established requirement of ex-post evaluation.

This new situation raises a number of fundamental challenges for ex-ante impact assessment and subsequent impact evaluation, because new mission-oriented programs show a number of specific characteristics, such as the need for policy mixes, which turn them into "systemic policies in a nutshell".

Recently, a conceptual and methodological approach has been developed which provides a frame of reference for impact assessment and impact evaluation of new mission-oriented programs (Weber and Polt 2014), but the empirical validation has been missing so far. This paper aims at revisiting some recent of the Austrian, European and OECD experiences, and at extracting lessons on the potential and the limitations of assessing and evaluating new mission-oriented programs. The analysis points to some converging insights across these cases, but also to different levels of aspiration that assessments/evaluations could strive for.

KEY FEATURES OF NEW MISSION-ORIENTED R&D PROGRAMS

New mission-oriented R&D programs show a number of specific features that need to be taken into account in their impact assessment and evaluation (see Foray et al. 2012; Dachs et al. 2015):

- Most recent MOPs – corresponding to the nature of societal challenges – are addressing issues that are broader in nature and scope than earlier technology-centered variants of MOPs. They involve a multitude of actors and stakeholder and deal with much longer time-horizons. This has considerable bearing on the role and weight of public and private actors, but also of other stakeholders.
- It has also become a frequently used design feature of MOPs that they span from basic research all the way through diffusion and implementation, hence the whole innovation (policy) cycle. This is because the ambition of MOPs is not just to foster innovation, but to trigger processes of socio-technical change that require the diffusion of the innovations in question, as well as wider systemic changes to happen.
- This in turn requires the coherent use of a substantial number of the instruments available in the toolbox of RTI policy and beyond, ranging from programs stimulating (oriented) basic research to the development of business models which would foster a rapid up-take of the respective technology. Especially demand-side instruments come into play here, as well as sectoral or thematic policies in key areas such as energy, health, agriculture, or environment. The choice of the appropriate 'policy mix' might again differ between the areas (e.g. aging societies, food-safety, climate change etc.)
- In the same vein, the goals and objectives of MOPs have become diverse. In contrast to single-issue programs like the often-cited role model of the earlier types of MOPs (e.g. the Manhattan and the Apollo programs) even programs confined to one field or topic (e.g. the US energy programs) are expected

to serve multiple goals, ranging from the mission in the narrow sense to commercial effects at the level of the individual participating firm to effects on other policy areas like national security and the like.

REQUIREMENTS AND CHALLENGES OF IMPACT ASSESSMENT AND EVALUATION OF MOPS

These characteristic features of MOPs point to some important challenges for impact assessment and evaluation. First, while typical economic micro-level effects can be analyzed with the help of well-established assessment and evaluation methods, this systemic policy approach typical of MOPs poses considerable challenges for the assessment of impacts with regard to higher-order mission goals: First of all, the impact of MOPs has to pass through different stages before it can actually exert an influence on new mission goals. The immediate impact of a mission-oriented R&I program occurs at the level of the participating firms or research organizations, where new research results are produced and – at least in some cases – innovations are introduced to the market. However, it is only after widespread adoption and diffusion of an innovation in the target system that an impact of a mission-oriented R&I program on higher-order mission goals can be observed. In several cases of MOPs, far-reaching transformative changes in the target system are needed to realize mission goals; changes that can at best be triggered and facilitated by research and innovation.

Secondly, for mission goals to be realized, complementary changes are also needed at different levels of the target systems. Borrowing from the multi-level perspective on socio-technical transitions (Geels 2005), change processes in technological niches and for individual firms (micro-level) can be distinguished from shifts in the socio-technical regimes (meso-level), and possibly even at the level of socio-technical landscape (macro-level). Most "new missions" funding programs tend to be defined at the level of such meso-level socio-technical regimes. Realizing these missions requires the widespread adoption and diffusion of innovations, including a transformation of production and consumption practices. Mission-orientation thus enhances a well-known problem of impact assessment, namely the attribution problem. These programs thus call for a different approach to impact attribution than single-target/single instrument programs.

Third, given the long term time horizon until the impacts of MOPs on mission goals materialize, and the uncertainty associated to both goals and impacts, adaptation and learning need to become an integral part of design and implementation of MOPs. The insights from a continuous monitoring of systemic effects of innovation and diffusion at different levels, as well as any improvements in the understanding of the mission need to be fed back into the program.

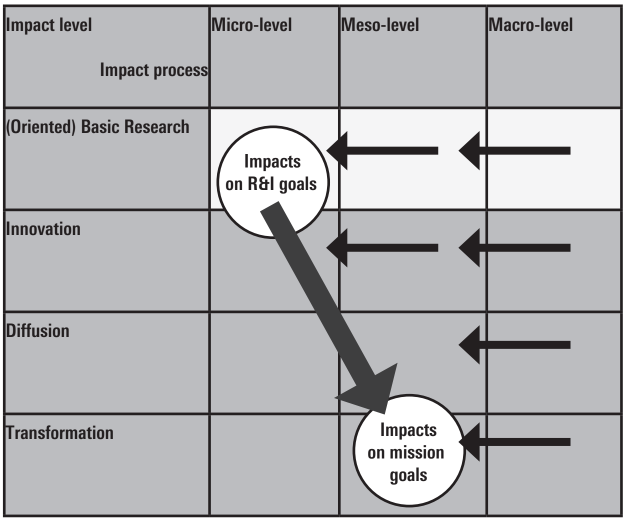
These challenges ask for a different approach to evaluation (see Mazzucato 2015), a requirement which has been increasingly recognized, but rarely put into practice.

AN INTEGRATED IMPACT ASSESSMENT APPROACH FOR NEW MISSION-ORIENTED PROGRAMS

In order to guide the assessment and evaluation of MOPs, we have suggested a conceptual and methodological framework. Conceptually, it points to different impacts levels and impact processes to be considered (see Figure 1):[1]

- **'Impact processes'**: Impact pathways range from thematically oriented, sometimes even basic, research to innovation, diffusion and system transformations, with the latter two stages being particularly relevant to new missions goals. At the earlier stages, RD&I funding directly affects the realization of research and innovation activities in firms and research organizations, i.e. at micro-level. Here, impacts can be measured rather directly (though not always comprehensively). At the later stages, at which mission targets are usually defined, effects only materialize to the extent that the innovations can be taken up (diffusion) and transformative processes are induced.
- **'Impact levels'**: Contributing to the achievement of mission goals implies changes to be realized at different levels, i.e. changes at micro-level of individual behavior, as well as at meso-level of structures and institutions, which in turn are embedded in change processes at macro-level. In some cases, the transformative processes may also affect this wider macro-level.

Figure 1: Conceptual framework of impact processes and impact levels



Methodologically, it suggests a scenario-based approach to ex-ante impact assessment, in order to take into account the openness, uncertainty and contingency on systemic changes of future impacts. Ex-post impact evaluation depends to a large extent on the framing of ex-ante impact assessment, in terms of 'tracing back' (most likely in a case study manner) specific impulses that were in the end strong enough to change the system (e.g. by being able to identify for the effects of the results from basic research to the achievement from mission-oriented research). In doing so, ex-post assessment would be a source for general 'policy learning', e.g. about the respective roles of basic research, social and institutional change and other dimensions that can drive systems change.

1 See Weber and Polt (2014) for further details on the impact assessment approach for MOPs, which we have labelled PESCA (Prospective & Adaptive Societal Challenges Assessment).

ANALYZING EXPERIENCES WITH THE MOP ASSESSMENTS AND EVALUATIONS USING THE PESCA FRAMEWORK

Over the past few years, some first impact assessments and evaluations of MOPs have been conducted. In this paper, we will look into selected a sample of experiences from national policy contexts (e.g. Environmental Impact Assessment of Austrian Technology Programs, Joint Programming Initiatives like FACCE or Urban Europe), European policy (e.g. the consecutive Impact Assessments of FP 7 and Horizon 2020) and OECD (recent work on impact assessment, applied to the evaluation of system innovation policies), all aiming to (i) establish a rationale for mission-oriented policies, (ii) in doing so, discussing a framework for the assessment of these policies or (iii) even try to assess or evaluate MOPs, in order to study the possibilities and limitations of assessing and evaluating MOPs.

We will analyze these examples against the background of the (elements of the) PESCA approach in order to find out whether the approach we had proposed is (at least partially) put into practice, what the experiences are with the bits and pieces of the approach and into which barriers and pitfalls the respective assessment processes have encountered.

These first experiences show that while there is still a lot of experimentation going on with assessing and evaluating MOPs, two main directions of work can be distinguished. On the one hand, efforts have been made to fully embrace the challenge of exploring and analyzing long-term system-level impacts using scenarios and modelling techniques. Often, the more narrowly defined (and more easily measurable) innovation-related impacts are distinguished from the more far-reaching systemic impacts (using a mix of qualitative and quantitative impacts scenarios or narratives). This direction could be called "comprehensive" impact assessment/evaluation.

On the other hand, a more "evolutionary" approach of improving existing empirical indicator frameworks can be observed, aiming to trace the different effects (from specific to systemic) that can be related to a MOP.

In the end, we will try to synthesize the findings of the cases into recommendations for further development of the PESCA approach.

REFERENCES

Dachs, B. et al. (2015): Herausforderungen und Perspektiven missionsorientierter Forschungs- und Innovationspolitik [Challenges and Perspectives of Mission-Oriented Research and Innovation Policy], Report to the German Expert Commission Research and Innovation, AIT/Fraunhofer ISI, Vienna/Karlsruhe

Foray, D., Mowery, D., Nelson, R. (2012): The need for a new generation of policy instruments to respond to the Grand Challenges research policy, Vol. 41, Issue 10, 1697-1792 (December 2012)

Geels, F. (2005): Technological transitions and system innovations: a co-evolutionary and socio-technical analysis. Cheltenham: Edward Elgar.

Link, A.N.; Scott, J. (2011): Public Goods, Public Gains: Calculating the Social Benefits of Public R&D, Oxford University Press, 2011.

Magro, E., Wilson, J.R. (2013): Complex innovation policy systems: towards an evaluation mix. In: Research Policy, Vol. 42, pp 1647-1656

Mazzucato, M. (2015): From Market Fixing to Market-Creating: A New Framework for Economic Policy. University of Sussex Working Paper Series. SWPS 2015-25 (September)

OECD (2015): Scope of the Work on Impact Assessment. DSTI/STP/TIP (2015)7

Polt, W. (2015): Qualitative and quantitative approaches in assessing mission-oriented RTD policies. Presentation at the OECD Workshop on Impact Assessment: Practices, Techniques and Policy Challenges 15-16 May, Tallinn

Weber, K.M., Polt, W. (2014): Assessing mission-orientated R&D programs: combining foresight and evaluation, FT Eval Journal, 39(2014), 5-10

AUTHORS

K. MATTHIAS WEBER
AIT Austrian Institute of Technology
Innovation Systems Department
Austria

WOLFGANG POLT
Joanneum Research POLICIES
Austria

VARIOUS APPROACHES TO MEASURING BUSINESS INNOVATION: THEIR RELEVANCE FOR CAPTURING SOCIAL INNOVATION

ATTILA HAVAS

The proposed paper reviews business innovation indicators from theoretical and policy perspectives. It discusses two widely used sets of innovation indicators, their context and shortcomings with the main objective to consider if they can be followed as a 'model' when designing social innovation indicators.

The main findings can be summarised as follows. Various economics paradigms treat (business) innovation (if not neglect it altogether) in diametrically different ways: they consider different notions as crucial ones (e.g. risk vs. uncertainty, information vs. various forms, types and sources of knowledge, skills and learning capabilities and processes); offer diverse justifications (policy rationales) for policy interventions; interpret the significance of various types of inputs, efforts, and results differently, and thus – implicitly – identify different 'targets' for measurement, monitoring and analytical purposes (what phenomena, inputs, capacities, processes, outcomes and impacts are to be measured and assessed).

The science-push model of innovation, reinforced by the sophisticated – and thus appealing and compelling – models of mainstream economics emphasises the economic impacts of R&D-based innovation efforts, advances the market failure argument and the concomitant set of policy advice. Hence it focuses the attention of decision-makers and analysts to the so-called ST mode of innovation. Measurement and monitoring systems influenced by this way of thinking – most notably the Innovation Union Scoreboard of the European Commission, but to a significant extent several other attempts, too, e.g. the Global Innovation Index, and the Technology Achievement Index compiled for the 2001 edition of the Human Development Report – tend to pay attention mainly to the ST mode of innovation, at the expense of the so-called DUI mode of innovation. It is a major concern, however, as the latter one is equally important from the point of view of enhancing productivity, creating jobs and improving competitiveness.

In contrast, evolutionary economics of innovation – in line with the networked model of innovation – stresses the systemic nature of innovation and thus advocates rectifying any systemic failure that hinders the generation, circulation and exploitation of any type of knowledge required for successful innovation processes. This way of thinking has influenced the measurement and monitoring practices of the European Commission or the OECD to a significantly lesser extent than mainstream economics.

In sum, the IUS indicators in principle could be useful in settings where the dominant mode of innovation is the ST mode. In practice, however, both the ST and DUI modes of innovation are fairly important. (Jensen et al., 2007) Moreover, the so-called Summary Innovation Index – calculated from the IUS indicators – does not provide sufficient information to assess a given innovation system: its low value could reflect either a low level of innovation activities altogether or a low level of R&D-based innovation activities (while other types of innovations are abundant). Yet, that is a fairly important distinction both from an

analytical and a practical (policy) point of view: these two innovation systems are fundamentally different. Analysts and policy-makers dealing with innovation, therefore, should pay attention to both R&D-based (ST) and non-R&D-based (DUI) innovations.

Further, while social innovations can certainly rely on R&D-based technological innovations, their essence tends to be organisational, managerial and behavioural changes. The IUS indicators do not capture these types of changes. More generally, analysts and decision-makers should be aware of the diversity of social innovations, too, in terms of their nature, drivers, objectives, actors, and process characteristics.

The Global Innovation Index (GII) has a significantly broader coverage – compared to the IUS – in two respects: it covers well over 100 countries, and considers 81 indicators, arranged in 7 "pillars". The seven pillars used in the 2014 edition of the GII include: Institutions (9 indicators), Human capital and research (11), Infrastructure (10), Market sophistication (10), Business sophistication (14), Knowledge and technology outputs (14), and Creative outputs (13). Concerning the composition of these pillars, a few observations are highlighted below.

Not all the elements considered in Pillar 1 are institutions ("rules of the game"), and not all are directly related to innovation processes and performance. It can be argued, though, that the aspects (attempted to be) captured by these indices are relevant to characterise the political, regulatory and business environment for innovation. Among the important missing elements, one should mention legislation on competition, as well as the entrepreneurial culture in a given country.

As for Pillar 2, its name is more 'ambitious' than its actual content. Life-long learning and other, informal modes of learning are also important factors, but not covered at all. While research is conducted outside universities, too, both by other publicly financed research organisations and businesses, these processes are not considered. Finally, university rankings are taken at face value, although these suffer from several major methodological weaknesses.

There is a certain mismatch between the name of Pillar 3 and its actual content, too.

The first sub-pillar of Pillar 6 is meant to be composed of indicators on "the result of inventive and innovative activities". Yet, most of these indicators are relevant to characterise R&D (and not innovation) activities. As for the knowledge impact sub-pillar, only one of the five components is related to knowledge impacts, and even that one is only partially: reflecting the impact of certain types of knowledge. As for knowledge diffusion, all the four components of that sub-pillar can indicate knowledge diffusion outside a given country (with certain limitations), and thus none of these seems to be relevant to characterise knowledge diffusion inside a given country.

In sum, the GII is a remarkable effort both in terms of its geographic and thematic coverage, but it suffers from severe weaknesses concerning business innovation activities. In several cases there is a non-negligible mismatch between the 'headline' notions (pillars and their sub-pillars) and the actual components (indices or indicators) selected. Just as in the case of the EIS and IUS indicators, there is a bias towards R&D-based (ST mode) innovations, and thus the DUI mode is eclipsed. It is even worse when R&D and innovation are conflated. As for describing and assessing social innovations, it would not be a fruitful effort to rely on any of the 81 GII indicators to describe and characterise social innovations.

The Technology Achievement Index, presented in the 2001 edition of the Human Development Report (UNDP, 2001) does not offer a promising approach, either. It is not a comprehensive measure: it considers only certain types of technological achievements and not necessarily those that are the most relevant from the point of view of human development. (Chiappero-Martinetti, 2015; see also Desai et al., 2002)

Some more general methodological lessons, however, can be distilled from the efforts devoted to measure business innovations. The first one concerns the use of composite indicators. Scoreboards and league tables compiled following the science-push logic, based on a composite indicator to establish rankings, and published by supranational organisations, can easily lead to 'lock-in' situations. National policy-makers – and politicians, in particular – are likely to pay much more attention to their country's position on a scoreboard than to nuanced assessments or policy recommendations in lengthy documents, and hence this inapt logic is 'diffused' and strengthened at the national level, too, preventing policy learning and devising appropriate policies. Despite the likely original intention, that is, to broaden the horizon of decision-makers by offering internationally comparable data, these scoreboards and league tables strengthen a narrow-minded, simplifying approach.

In other words, given the diversity among innovation systems, one should be very careful when trying to draw policy lessons from the 'rank' of a country as 'measured' by a composite indicator. A scoreboard can only be constructed by using the same set of indicators across all countries, and by applying an identical method to calculate the composite index. Yet, it is important to realise that poor performance signalled by a composite indicator, and leading to a low ranking on a certain scoreboard, does not automatically identify the area(s) necessitating the most urgent policy actions.

In contrast, a high ranking on a scoreboard, e.g. Sweden's first place on the 2013 Innovation Union Scoreboard does not necessarily reflect a satisfactory performance. By taking into account the input and output nature of various IUS indicators Edquist and Zabala- Iturriagagoitia (2015) calculated the productivity of national innovation systems covered by the IUS and using this assessment – which is, no doubt, highly relevant form a policy point of view – Sweden ranks a mere 24.

Analysts and policy-makers, therefore, need to avoid the trap of paying too much attention to simplifying ranking exercises. Instead, it is of utmost importance to conduct detailed, thorough comparative analyses, identifying the reasons for a disappointing performance, as well as the sources of – opportunities for – balanced, and sustainable, socio- economic development.

Second, the degree of novelty and the unit of analysis are interrelated issues when business innovations are surveyed. It looks a rather difficult task to establish the degree of novelty of a given social innovation. Actually, this issue seems to be of lesser importance in these cases: intellectual property rights are seldom an issue for social innovators. Prestige – obtained by being acknowledged as a creative social innovator – might, however, play a role: it could be perceived as an incentive to initiate social innovation projects. No doubt, it is an empirical question to establish the role of prestige in these endeavours.

It could be also an interesting – but certainly a demanding – research question to identify whether a given social innovation is a standalone new solution or – using the analogy of technology systems – a part of a new 'social system', that is, a set of socially, institutionally, organisationally, and economically interconnected social innovations, affecting several groups of people or an entire community (a neighbourhood, village, town or city) at the same time, occasionally leading to the emergence of new social structures, norms, institutions, behaviour, value systems and practices at a higher level of aggregation (e.g. sub-national regions, nations or even supra-national regions [for example, the European Union]).

Efforts aimed at measuring social innovation cannot rely on a long-established tradition. The proposed TEPSIE framework for measuring social innovation (Bund et al., 2013) has been a significant effort to this end, but it needs some non-negligible improvement. Its first pillar, called entrepreneurial activity is not specific to social innovation, on the one hand, and somewhat neglects non-entrepreneurial social innovation activities, on the other. Its second pillar, called field-specific output and outcomes, offers useful hints, but we are faced by the usual attribution problem in the case of social innovations, too. The third pillar is concerned with framework conditions. The structure of the TEPSIE indicators prompts a more general caveat: analysts and policy-makers need to be aware of the differences between measuring (a) social innovation activities (efforts) themselves; (b) the framework conditions (pre-requisites, available inputs, skills, norms, values, behavioural patterns, etc.) of being socially innovative; and (c) the economic, societal or environmental impacts of social innovations.

SELECTED REFERENCES

Bender, G., D. Jacobson, P.L. Robertson (eds) (2005): Non-Research-Intensive Industries in the Knowledge Economy, Perspectives on Economic and Social Integration – Journal for Mental Changes, special issue, XI (1–2)

Bund, E., D-K. Hubrich, B. Schmitz, G. Mildenberger, G. Krlev (2013): Blueprint of social innovation metrics – contributions to an understanding of opportunities and challenges of social innovation measurement, deliverable 2.4 of the TEPSIE project: The theoretical, empirical and policy foundations for building social innovation in Europe, European Commission 7th Framework Programme

Bush, V. (1945): Science: the Endless Frontier, Washington DC: US Government Printing Office

Caraça, J., B-Å. Lundvall, S. Mendonça (2009): The changing role of science in the innovation process: From Queen to Cinderella?, Technological Forecasting and Social Change, 76 (6): 861–867

Castellacci, F. (2008): Innovation and the competitiveness of industries: Comparing the mainstream and the evolutionary approaches, Technological Forecasting and Social Change, 75 (7): 984–1006

Chiappero-Martinetti, E. (2015): Relationship between innovation/ technology and human development, contribution to WP3 of the CrESSI project, Pavia: mimeo

Cornell University, INSEAD and WIPO (2014): The Global Innovation Index 2014: The Human Factor in Innovation, Fontainebleau, Ithaca, and Geneva: Cornell University, INSEAD and WIPO

Desai, M., S. Fukuda-Parr, J. Johansson, F. Sagasti, (2002): Measuring Technology Achievement of Nations and the Capacity to Participate in the Network Age, Journal of Human Development, 3 (1): 95–122

EC (2015): Innovation Union Scoreboard 2014, Directorate-General for Internal Market, Industry, Entrepreneurship and SMEs, Brussels: European Commission

Freeman, C. (1994): The economics of technical change: A critical survey, Cambridge Journal of Economics, 18 (5): 463–514

Godin, B. (2008): The moral economy of technology indicators, in: Hirsch-Kreinsen, H. and D. Jacobson (eds) (2008), pp. 64–84

Grupp, H. and T. Schubert (2010): Review and new evidence on composite innovation indicators for evaluating national performance, Research Policy, 39 (1): 67–78

Fagerberg, J., D.C. Mowery, R.R. Nelson (eds) (2005): The Oxford Handbook of Innovation, Oxford: Oxford University Press

Hall, B.H. and N. Rosenberg (eds) (2010): Economics of Innovation, Amsterdam: North- Holland

Hatzichronoglou, T. (1997): Revision of the High-Technology Sector and Product Classification, OECD STI Working Papers, 1997/2

Hirsch-Kreinsen, H., D. Jacobson, S. Laestadius (eds) (2005): Low Tech Innovation in the Knowledge Economy, Frankfurt: Peter Lang

Hirsch-Kreinsen, H. and D. Jacobson (eds) (2008): Innovation in Low-Tech Firms and Industries, Cheltenham: Edward Elgar

Jensen, M.B., B. Johnson, E. Lorenz, B-Å. Lundvall (2007): Forms of knowledge and modes of innovation, Research Policy, 36 (5): 680–693

Kline, S.J. and N. Rosenberg (1986): An Overview of Innovation, in: Landau, R., Rosenberg, N. (eds): The Positive Sum Strategy: Harnessing Technology for Economic Growth, Washington: National Academy Press, pp. 275–305

Laestadious, S., T.E. Pedersen, T. Sandven (2005): Towards a new understanding of innovativeness – and of innovation based indicators, in: Bender et al. (eds) (2005), pp. 75– 121

Nelson, R.R., S.G. Winter, 2002, Evolutionary Theorizing in Economics, Journal of Economic Perspectives, 16 (2), 23–46

Pavitt, K. (1999): Technology, Management and Systems of Innovation, Cheltenham: Edward Elgar

Robertson, P. and K. Smith (2008): Distributed knowledge bases in low- and medium technology industries, in: Hirsch-Kreinsen, H., Jacobson, D. (eds) (2008), pp. 93–117

Robertson, P., K. Smith, N. von Tunzelmann (2009): Innovation in low- and medium- technology industries, Research Policy, 38 (3): 441–446

Smith, K. (2005): Measuring Innovation, in: Fagerberg et al. (eds) (2005), pp. 148–177 UNDP (2001): Human Development Report: Making new technologies work for human development, Oxford: Oxford University Press von Tunzelmann, N. and V. Acha (2005): Innovation in "Low-Tech" Industries, in: Fagerberg et al. (eds) (2005), pp. 407–432

AUTHOR

ATTILA HAVAS
Institute of Economics, CERS, HAS
1112 Budapest, Budaörsi út 45.
Hungary
phone: +36308164266

EXPLAINING THE SUCCESS (AND FAILURE) OF THE INTERVENTION WITH THE USE OF SAMPLING BASED ON PROPENSITY SCORE MATCHING

MACIEJ KONIEWSKI, SEWERYN KRUPNIK AND PAULINA SKÓRSKA

BACKGROUND

Combining theory-based and counterfactual approaches is perceived by many experts (White, 2009) and stakeholders (European Commission, 2014) as the best approach for conducting ex-post evaluations. However, evaluators each time have to design the research which would use the valid methodology – corresponding to the research questions and available resources. Thus, there is need for the diverse research schemes which would combine both approaches.

AIM

The article presents the unique approach to the explanation of success (and failure) of the specific public interventions. It uses sampling based on propensity score matching (PSM). The approach was applied within the evaluation of financial support received from Innovative Economy Operational Programme (IE OP) in Poland. It was applied to the measure 4.4 IE OP within which investment projects involving the purchase or implementation of research results/new technological solutions were supported.

In the study two counterfactual approaches were used. Firstly, the traditional counterfactual approach allowing estimation of the net effects of selected measures of IE OP (comparing situation of beneficiary after receiving support to the situation in which intervention would not been implemented). Secondly, modified counterfactual approach was used to enable estimation of the relative causal effects of the selected measures (comparing situation of the beneficiary of the measure to a situation in which it would be the beneficiary of another measure).

The analysis of effects for 4.4 IE OP indicated probable achievement of the objectives related to R&D (Research and Development). Beneficiaries of the measure at the end of the support more often than non-beneficiaries experienced internal and external expenditures on R&D activities. Due to the much higher value of funding obtained under measure 4.4 than under other measures, one could expect larger effects for the beneficiaries of this measure in almost all of the analyzed categories of effects (financial performance, export, innovation, R&D). The results of the analysis did not confirm these expectations. For most economic indicators there was no effect - no difference between the situations of the analyzed groups of beneficiaries (measure 4.4 versus other measures). It is also worth to mention the relatively higher percentage of beneficiaries of measure 4.4 showing a loss as compared to the beneficiaries of measure 4.2 (Fig. 1).

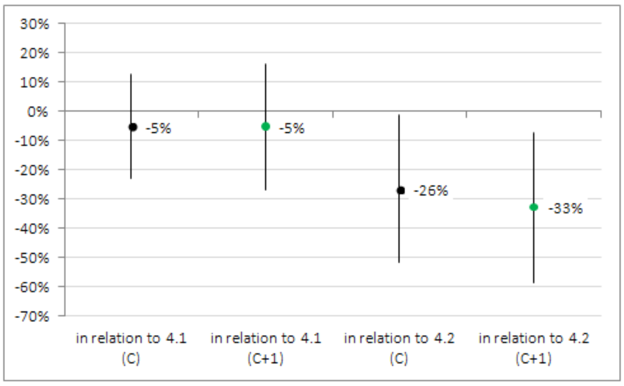


Figure 1. Proportion of 4.4. IE OP beneficiaries not indicating a loss (C – year of project completion; C+1 – one year after project completion). The graph presents the point estimates (Difference - In - Difference for trimmed means) with confidence intervals. If the confidence interval (horizontal line) crosses the y-axis at the value of zero, the effect is not statistically significant. If it does not cross y-axis at 0, the effect is statistically significant. Colors differentiate effects: black dots were used to indicate the difference (effect) between the start and the year of completion of the project (C), and the green dots between a start and year after completion of the project (C+1).
n: 4.1=89, 4.2=72, 4.4=140.

These results are surprising in the context of a higher funding received by the beneficiaries of measure 4.4. Thus, there was a need to explain why, on average, the financial situation of measure 4.4 beneficiaries after support was not as good as expected.

METHOD

Out of program beneficiaries, the contrast pairs were selected to in-depth interviewing. The pairs' selection procedure was as follows:

1. The program success criterion was defined at first using the modified Return On Sales (ROS) indicator computed as gross profit divided by net sales revenue, for companies which reported no loss for the fiscal year previous to the measurement point.
2. The linear regression model was computed for ROS declared for one year after a company had completed the project funded within the program as a dependent variable and ROS for the year when the company started the project and set of company and received funding details as independent variables, which were: year of the project launch, year of the project completion, the company type (Ltd. vs. stock), the company size (up to 50 employees, 50-250, over 250), sector (production vs. other), year in which the company was established, percent of own input in total project costs besides funding received, amount of founding received, voivodeships (region) of the company headquarters). The model proved to have satisfactory diagnostic results with 23% variance explained (the adjusted R-squared). The purpose of the model was to capture as many available information in dataset as possible.
3. From the model the residuals were obtained, which can be interpreted as a measure of difference from expected ROS level one year after project completion. Therefore, the residuals were interpreted as a success indicator: either a company made a good use of funding received or not. This logic of the residuals interpretation come from Value Added models (Hibpshman, 2004), which are popular, e.g. in assessment programs of teachers effectiveness.
4. The companies were ranked ordered descending based on residual values. Companies from above the third quartile and below the median were kept for further analysis constituting two groups: successful companies and companies which failed to turn received funding into ROS increase.
5. The Propensity Score Matching (PSM) - using the same covariates as in linear regression model described above - was performed to find the most alike pairs of companies which differ only in being classified as those, which succeed or failed to turn received funding into ROS increase.
6. For each successful company two failing companies (controls) were matched if possible, hence some triplets were matched.

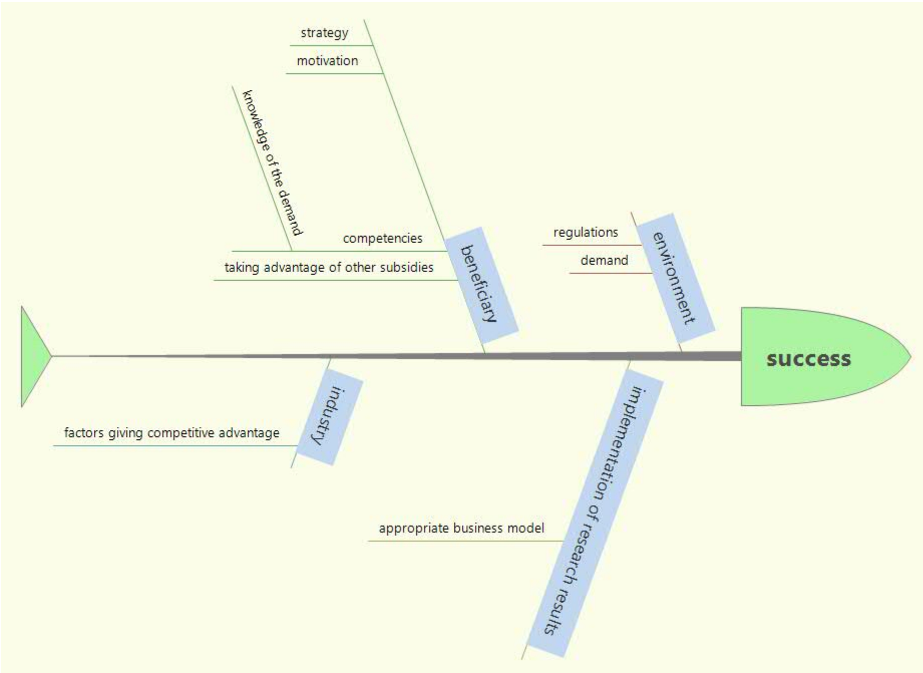
It was expected that this approach will increase chances of selecting for in-depth interviews companies which allow highlighting program success and failure factors. Then, 16 in-depth interviews with beneficiaries (8 successful and 8 unsuccessful) were conducted.

RESULTS

In the qualitative part of the study we identified a number of factors affecting the translation of the results of the project to changes in profitability (Fig. 2). These factors are linked to the characteristics of the beneficiaries (e.g. competencies of the managers, strategy, motivation, taking advantage of other subsidies), the specificity of the industry sector (factors giving competitive advantage), way of implementing the projects (appropriate business model), as well as socio-economic environment (demand and business regulations).

Figure 2. Factors accounting for success or failure of translating the effects of supported projects into the increase of the profitability.

The following key factors led to the situation in which some entrepreneurs experienced a relatively small effect on the profitability of the company:



1. general economic situation (economic crisis) and difficulty in prediction of changes in the industry (e.g. energy prices, the embargo on food products in Russia);
2. ack of adequate monitoring of the market situation in terms of demand and appropriate plan to reach customers[1].

CONCLUSIONS

Thanks to the application of new, described approach the factors accounting for the lack of the success of the intervention were identified. Thus, the approach proved to be useful in explaining the surprising results of counterfactual analysis. It was recommended that the identified

1 The full report (with English abstract) is available online. https://badania.parp.gov.pl/images/badania/Raport_Barometr_netto_POIG.pdf

factors of success and failure should be taken into account in the process of project selection in the analogous measures in next Operational Programme Smart Growth under the programming period 2014 -2020.

What was both surprising and confirming the usefulness of the approach was the compatibility of the objective classification based on ROS with the subjective perception of success articulated by beneficiaries within in-depth interviews. Beneficiaries, whose projects were classified as successful, stated that the projects have had great impact on the growth of their companies. And beneficiaries, whose projects were classified as unsuccessful, admitted that the results of the projects did not match their expectations. The result is surprising because it is uncommon for beneficiaries of the public support to make such statements.

While the approach provides robust results and conclusion it could be further enhanced by the application of qualitative comparative analysis (QCA). QCA enables investigation of interactions between factors influencing effects and robust procedures for identification of key factors (Befani, 2013).

REFERENCES

Befani, B. (2013): "Between complexity and generalization: addressing evaluation challenges with QCA", Evaluation, vol. 19, pp. 269-283.

Hibpshman, T.L. (2004): A review of value-added models. Kentucky Education Professional Standards Board.

Koniewski, M., S. Krupnik, P. Skórska, K. Turek and T. Geodecki (2016): "Analiza efektów netto wybranych działaĐ Programu Operacyjnego Innowacyjna Gospodarka 2007-2013, z wykorzystaniem podejĐcia counterfactual impact evaluation w ramach projektu ewaluacyjnego PARP „Barometr InnowacyjnoĐci"", retrieved on 13 April 2016 from https://badania.parp.gov.pl/images/badania/Raport_Barometr_netto_POIG.pdf

White, H. (2009): "Theory-Based impact evaluation: principles and practice", retrieved on 20 March 2016 from http://www.3ieimpact.org/media/filer_public/2012/05/07/Working_Paper_3.pdf.

European Commission (2014): "The programming period 2014 - 2020. Guidance document on monitoring and evaluation, European Cohesion Fund, European Regional Development Fund. Concepts and recommendations", retrieved on 10 March 2016 from http://ec.europa.eu/regional_policy/sources/docoffic/2014/working/wd_2014_en.pdf

AUTHORS

MACIEJ KONIEWSKI
Center for Evaluation and Analysis of Public Policies
Jagiellonian University, Philosophy Department,
Gołębia St. 24, 31-007 Kraków,
Poland

SEWERYN KRUPNIK
Corresponding author
Center for Evaluation and Analysis of Public Policies
Jagiellonian University, Philosophy Department,
Gołębia St. 24, 31-007 Kraków,
Poland
+48 12 663 17 92

PAULINA SKÓRSKA
Center for Evaluation and Analysis of Public Policies
Jagiellonian University, Philosophy Department,
Gołębia St. 24, 31-007 Kraków,
Poland

EX-ANTE EVALUATION OF RESEARCH POLICY: AN AGENT-BASED MODEL OF AUSTRIAN BIOTECHNOLOGY

MANFRED PAIER, MARTINA DUENSER, THOMAS SCHERNGELL AND SIMON MARTIN

1. MOTIVATION

Over the recent past, we can observe increasing interest in the ex-ante impact assessment[1] (IA) of public research policy (Delanghe and Muldur 2007), mainly related to the growing importance of accountability and limited budgets. Rising demand and standards require flexible methods that go beyond extrapolations of current trends.

Hence, we propose an empirical agent-based model (ABM[2]) of knowledge creation in a system of interacting research firms to analyse the effects of policy interventions on the knowledge-related system output. Hereby, we combine an elaborate empirical initialisation and calibration strategy and econometric techniques using patent and company data. Following a quite recent development of empirical agent-based modelling (Smajgl and Barreteau 2014) the integration of empirics enables us to apply our model in real world contexts, such as in our case the area of research policy.

We contribute to the state of the art in two major respects: (i) the development of an elaborate empirical calibration strategy for the application of ABMs in innovation economics, especially concerning knowledge-driven industries, and (ii) the use of ABMs to support decision makers in research policy in the context of ex-ante impact assessment.

In our illustrative example, we focus on knowledge creation in the Austrian biotechnology sector. Biotechnology is a knowledge-driven industry, characterised by high intensity of both research and knowledge exchange among economic actors (Owen-Smith and Powell 2004, Tödtling and Trippl 2007). In this context, economic performance is to a substantial degree driven by the creation of knowledge and its direct application to the industrial context.

2. THE MODEL AND ITS APPLICATION TO BIOTECHNOLOGY

The basic structure of our model can be divided into an input side, with the agent's knowledge endowment and strategies, and an interaction part with research processes in order to generate knowledge and an output side, where the knowledge output is realised in terms of knowledge gains and patents.

However, the relationship between these three parts is not linear but is characterised by feedback-loops and interdependencies. The process of knowledge creation as a whole is embedded in a research system with sector specific institutional characteristics, which is by itself affected by research policy interventions at various stages in the model. An overview of the conceptual model is shown in Figure 1.

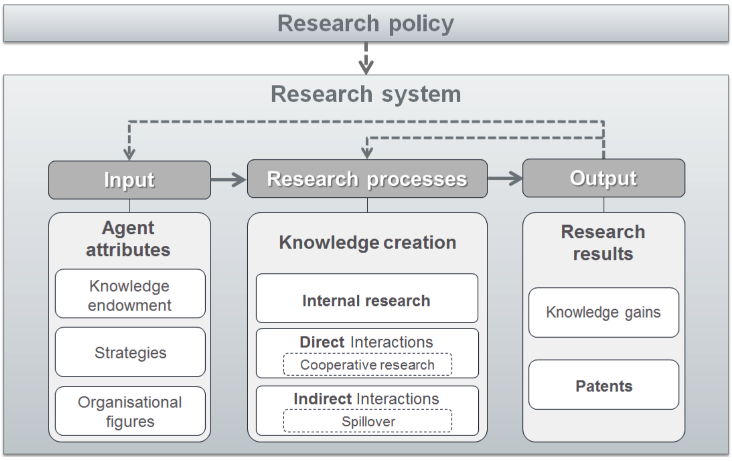


Figure 1: Conceptual model design

1 IA refers to the evaluation of the potential of a project or program to deliver benefits from proposed policy interventions and is – since his official introduction into policy making by the European Commission in 2002 – serving as a formal procedure to analyse potential effects of new policies like the European Framework Programmes before their implementation.

2 We use the acronym ABM to refer to both "agent-based-modelling" and "agent-based model".

2.1 AGENTS, ATTRIBUTES AND STRATEGIES

As agents in the model, we consider 61 industrial biotechnology firms located in Austria with active patenting records over the years 2000-2010. Each agent is provided with a knowledge endowment, strategies and characterising organisational figures, such that the agents are heterogeneous with respect to their attributes.

Knowledge Endowment. The knowledge endowment of an agent is defined as a set of three so-called kenes (Gilbert et al. 2001). Each kene consists of a technology class, a subfield associated with the respective technology class and an expertise level in the specific technology class and subfield. The technology classes are initialised according to the corresponding firm's patent portfolio. The subfield and the expertise level are random numbers between one and ten.

Strategies. Each time step (i.e. a quarter of a year), an agent engages in research activities comprising two distinct phases: (i) definition of a research target (ii) definition of a research strategy determining how to obtain this target. For (i), four alternative search strategies are possible: *gridlock* (no research is conducted), *conservative* (increasing expertise in currents research area), *incremental* (diversification to a new subfield) and *radical* (diversification to a new technology class). For the case of (ii) there are three kinds of research strategies: *spillover, internal research and cooperative research* (see subsection 2.2).

Organisational figures. Each agent is also individually equipped with four empirically based organisational figures: (i) research expenditures (ii) number of employees, (iii) assets and (iv) age, taken from a company database and a recent sector study (Schibany et al. 2010, Bureau van Dijk 2014).

2.2 AGENTS' PROCESSES

The research process starts with a certain probability that the agents may receive knowledge through spillover. If an agent finds an appropriate kene for matching its research target during the spillover process, the research result is taken for granted, which completes the research process. With the complementary probability, the agents engage with specified fractions either in cooperative research or in internal research in the first place. In case of a missing match during the spillover process the agents engage in research according to these fractions as well. The attainment of the research result depends on respective success rate parameters. If the research result is actually achieved through these research processes, the new kene replaces the old one in the knowledge endowment of the agent.

Whether the knowledge gains of an agent classify for becoming a patent, is determined by an empirical output filter (*fitness function*). The fitness function is composed of two parts: a system parameter and a function including the empirically estimated coefficients influencing the patenting propensity of an individual agent given its respective organisational figures. The coefficients are estimated by means of a zero-inflated negative binomial model (see Cameron and Trivedi 1998 for details) for a sample of 156 patenting and non-patenting Austrian biotechnology firms.

3. SCENARIO ANALYSIS

To assess the effects of different policy interventions with our model, we define scenarios[3] and analyse them through simulations[4]. Exemplary, the results of two funding alternatives are presented as scenarios with adapted parameter values in the model. As a reference, a baseline scenario is simulated with certain parameter values resulting from an initial model calibration. Note that the scenarios always refer to changes as compared with the baseline scenario.

SCENARIO: "FOCUS ON DIRECT VS INDIRECT FUNDING"

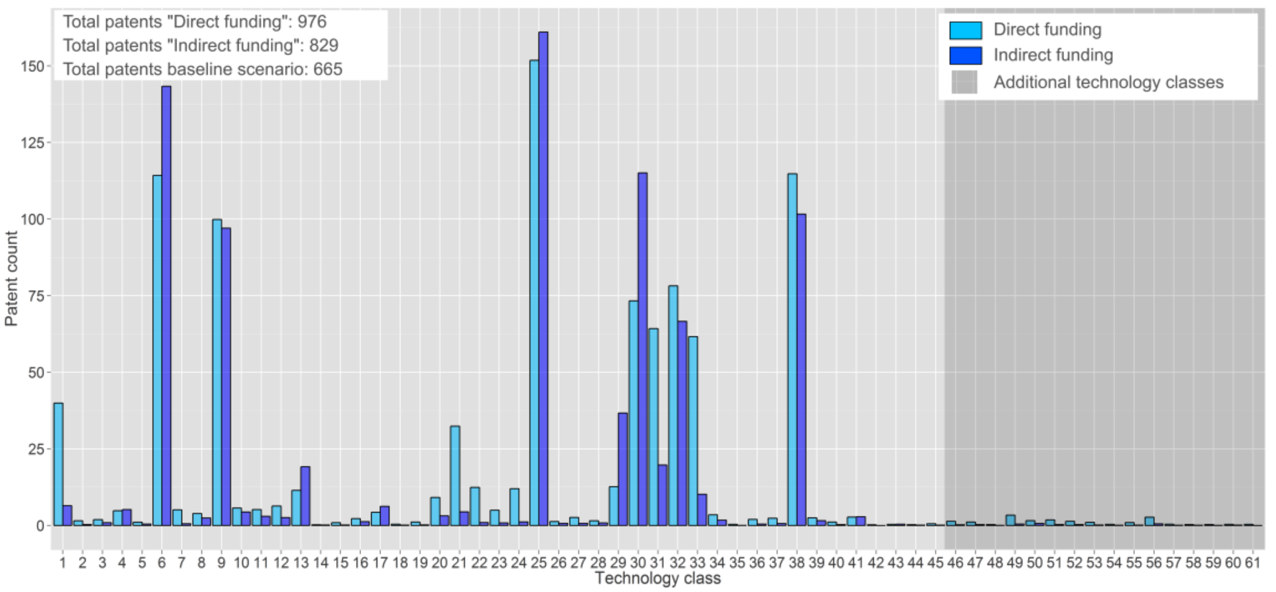
Direct funding subsumes the direct transfer of public financial resources to finance research projects. Due to the asserted substantial takeover of risk by the state, research agents may be encouraged to explore new areas of research. This might lower the entry costs of a firm or research organisation, building up new areas of expertise. It is important to note that direct funding is not only technology-specific or mission-oriented (with a steering effect on the development of new technology, "top-down funding"), but can also be independent of technology ("bottom-up funding") (Astor et al. 2009). In contrast to direct monetary grants, *indirect funding* takes the form of easing the tax burden if research was conducted, independent of the research area or research success. In this case the risk of engaging in research activities is also lowered to some degree the first place but the assumption is that risk reduction is not as pervasive as in the case of receiving subsidies via direct funding (Mohnen and Lokshin 2009).

In Figure 2 the simulation results of the scenario pair are illustrated. Regarding the total numbers of patents, both tested scenarios lie above the value of the baseline scenario; however, the patent count in the scenario of direct funding exceeds that of the indirect funding scenario. In the case of direct funding we observe a fairly strong diversification among the technology classes, i.e. direct funding especially promotes the "smaller" classes. This results from the increased number of radical agents and their expanded search horizon while choosing a new technology class as their research target. In contrast, indirect funding seems to only reach higher numbers of patents in a few large technology classes. This concentration on already predominant classes is due to the reduced share of agents with radical search strategy.

From an innovation policy perspective, these results are plausible. The scenario of indirect funds, for example, exhibits a characteristic phenomenon – windfall gains. In this case all agents are favoured, no matter if research would have been conducted anyway. This may indicate a reduced effectiveness of this funding type, with respect to diversification.

3 The simulations are conducted over a period of 30 years (i.e. 120 time steps). Furthermore, the results illustrated and discussed represent averages of 100 runs with varying random seeds.

4 The model is implemented in Java using the MASON platform (http://cs.gmu.edu/~eclab/projects/mason/) and the results are analysed with R.



Note: technology classes (T) with patents > 50: T 6 = Preperations for medical purposes, T 9 = Therapeutic activity of chemical compounds or medicinal preperations, T 25 = Peptides, T 30-33= Biochemistry, microbiology and enzymology, T 38 = Investigating or analysing materials by determining their chemical or physical properties.

Figure 2: Patents by technology class for direct vs. indirect funding (total after 120 steps)

4. CONCLUDING REMARKS

Through a transparent and robust model design, we aim at increasing the credibility of the ABM approach in the context of research policy. However, it has to be conceded that the complexity of the model presented here is so far mainly confined to knowledge creation, while economic aspects are deliberately kept simple. In this respect, future work will introduce additional complexity to the model with respect to exploitation of knowledge and population dynamics, in order to further increase its credibility step-by-step in the context of ex-ante impact assessment of policy intervention.

REFERENCES

Astor M, Heinrich S, Klose G, Riesenberg D (2009) Interventionslogik und Markt- / Systemversagen sowie Zusammenspiel der Institutionen und Akteure. In Systemevaluierung der österreichischen Forschungsförderung und -finanzierung. (ed.), Vol. Teilbericht 9, pp. Prognos AG, Berlin

Bureau van Dijk (2014) ORBIS company database. (ed.), Vol. 2014, pp.

Cameron AC, Trivedi PK (1998) Regression analysis of count data. Cambridge University Press, Cambridge

Delanghe H, Muldur U (2007) Ex-ante impact assessment of research programmes: the experience of the European Union's 7th Framework Programme. Science and Public Policy 34(3):169-183

Gilbert N, Pyka A, Ahrweiler P (2001) Innovation Networks - A Simulation Approach. Journal of Artificial Societies and Social Simulation 4(3):8

Mohnen P, Lokshin B (2009) What Does it Take for an R&D Tax Incentive Policy to Be Effective? . URL: http://d.repec.org/n?u=RePEc:cir:cirwor:2009s-11&r=ino CIRANO Scientific Series, Montreal

Owen-Smith J, Powell WW (2004) Knowledge Networks as Channels and Conduits: The Effects of Spillovers in the Boston Biotechnology Community. Organization Science 15(1):5-21

Schibany A, Berger M, Buchinger E, Dachs B, Dinges M, Ecker B, Falk M, Gassler H, Heller-Schuh B, Hofer R, Huber P, Janger J, Reinstaller A, Streicher G, Unterlass F (2010) Österreichischer Forschungs- und Technologiebericht 2010. In Lagebericht gem. § 8 (1) FOG über die aus Bundesmitteln geförderte Forschung, Technologie und Innovation in Österreich. (ed.), Vol. pp. BMWF, BMVIT, BMWFJ, Wien

Smajgl A, Barreteau O (Eds) (2014) The Characterisation and Parameterisation of Empirical Agent-Based Models. Empirical Agent-Based Modelling - Challenges and Solutions, Springer, New York Heidelberg Dordrecht London

Tödtling F, Trippl M (2007) Knowledge Links in High-Technology Industries: Markets, Networks, or Milieu? The Case of the Vienna Biotechnology Cluster. International Journal of Entrepreneurship and Innovation Management 2/3/4/5:345-365

AUTHORS

MANFRED PAIER
Innovation Systems Department, AIT Austrian Institute of Technology GmbH
Donau-City-Straße 1, 1220 Vienna
Austria
E manfred.paier@ait.ac.at
T +43 50550 4568

MARTINA DUENSER
Innovation Systems Department, AIT Austrian Institute of Technology GmbH

THOMAS SCHERNGELL
Innovation Systems Department, AIT Austrian Institute of Technology GmbH

SIMON MARTIN
Department of Economics, University of Vienna

ASSESSING THE IMPACT OF PUBLIC FUNDING AND TAX INCENTIVES IN RUSSIA: RECIPIENT ANALYSIS AND ADDITIONALITY EFFECTS EVALUATION

YURI SIMACHEV, MIKHAIL KUZYK AND NIKOLAY ZUDIN

In recent years the basic evaluation approach of the innovation policy toolbox has become the concept of additionality. Conceptually, in the context of government intervention the notion of additionality involves comparison of the real situation of receiving government support with a hypothetical scenario of what would have happened if no support had been provided. So far a considerable number of studies assessing the effects of innovation policy on firms' activity with the use of the concept of additionality has taken place. However, none of the known studies paid attention to the Russian innovation policy additionality.

The aim of our study is to perform a microeconomic evaluation of the industrial firms' support implementation in Russia focusing on its two main instruments: direct funding and tax incentives. The usage of these two instruments for the analysis is quite straightforward as they are traditionally viewed as the key elements of the national innovation policy toolbox (see, e.g., David et al., 2000; OECD, 2015) and are well ahead of the other instruments in their "coverage" – the number of the firms supported (Kuzyk, Simachev, 2013).

THE OBJECTIVES OF THIS STUDY ARE AS FOLLOWS:

- firstly, to identify the "typical profile" of the firms-beneficiaries of the government support policy as a whole and direct funding and tax incentives in particular;
- secondly, to consider the basic input, output and behavioral additionality effects;
- thirdly, to analyze the "relative" additionality of fiscal support and tax incentives.

Data was collected from a questionnaire survey of top executives of Russian manufacturing firms which was held in September-October 2015. The sample consists of 658 firms, ¾ of which belong to high-tech industries.

In order to identify the specifics of public support recipients and the achieved results frequency and regression analysis are used. Moreover, for a more precise definition of "relative" additionality effects of direct funding and tax incentives we use a propensity score matching (PSM) which is currently one of the main techniques for the analysis of the additionality at firm level (see, e.g., Fier et al., 2006; Baghana, 2010; Marzucchi, Montresor, 2013; Cantner., Kösters., 2015). An important distinguishing feature of our approach is that we analyze the additionality of a concrete instrument for a particular firm relative to all other used instruments and therefore consider the "relative" additionality. This enables us to highlight additionality effects inherent precisely to tax and financial instruments distinguishing them from the "background" of all other elements of the innovation policy toolbox.

THE KEY RESULTS OF THIS STUDY ARE:

1. Despite the fact that Russian industrial innovation policy toolbox is rather diversified there is a strong emphasis on the development of sufficiently large and long-operating companies. Such result is not surprising, especially for the Russian economy. Positive relationship between the size of the firms and the likelihood of receiving government support has been identified in a number of empirical studies (see, e.g., Fier, Heneric, 2005; Aschhoff, 2010; Simachev et al., 2014a). The question considering the relative efficiency of the government support of small and large firms is rather controversial. Today there exists empirical evidence of both significant influence of government support on SMEs, including behavioral changes (Loof, Heshmati, 2005; Wanzenbock et al., 2013), and substantial corresponding changes in the large firms (Falk 2006). Obtained results confirm, rather, the second point of view. However, due to the relatively small number of relevant observations, we can only hypothesize that in Russia instruments of government support (especially tax incentives) provide positive changes mainly for middle and large sized firms. Our view is that of the largest significance in the implementation of the instruments of government support are not the formal characteristics of the beneficiaries (such as size, age etc.) but their "quality". The recipients of government support should have big potential for further successful development and, what is more important, demonstrate the abilities to implement it. However, in Russian realities that principle is not always followed. In periods of relative economic stability the government mostly support successfully developing firms (see, e.g., Simachev et al., 2014a), whereas crises force the government to shift the support focus towards troubled companies, especially if these are of a great importance in the context of providing socio-economic stability in the region or/and in the whole country (Higher School of Economics, Interdepartmental Analytical Center, 2009; Mau, 2010).

2. Tax and financial instruments of the government support de facto have differential target audiences: the use of tax incentives is not likely for small firms, whereas medium-sized companies relatively rare appear to be the recipients of the financial support. The former can be the reflection of both the imperfect parameters of the tax instruments (their rate, base, etc.) for small businesses and the existence of significant implementation and administration problems, which are acceptable for large companies but excessive for small firms. As for the fact of a relatively rare financial support of the medium firms, it can be considered as another empirical evidence of a lack of instruments aimed at funding medium-sized projects and companies. (see also Simachev et al., 2012).
3. The relatively small impact of government support on science-business cooperation seems to us quite unexpected (abroad, this effect is among most frequently observed, especially in the case of financial support – see e.g. Busom, Fernandez Ribas, 2008; Idea Consult, 2009; Marzucchi, Montresor, 2013). Also this fact is rather discouraging, as the Russian government make considerable effort to enhance linkages and interactions between the R&D sector and industry. The absence of an explicit result of these efforts, to our mind, can be explained by the fact that government support often does not lead to the creation of new linkages and partnerships but only contributes to the "capitalization" of long-established ones (Simachev et al., 2014c). Note that a significant contribution of the government support to the improvement of existing science-business linkages and partnerships has been widely observed abroad (see, e.g., Georghiou et al., 2005; Lohmann, 2014).
4. Our empirical analysis as well as a significant number of earlier studies has confirmed the importance of the fiscal support in providing all main kinds of additionality. The main input effect is the increase of investment in new equipment; output - the increase of production of new and improved products, behavioral – the initiation of new perspective projects and an acceleration of project implementation. It should be noted that project additionality (government contribution to firms' launching new projects) is one of the most frequently observed behavioral changes (see, e.g., Falk, 2007; Idea Consult, 2009), what cannot be said about acceleration additionality (when government support speeds up the course of the project) which was analyzed by researchers to a considerably smaller extent.

Unlike financial instruments, tax incentives almost do not provide significant results in terms of additionality concept. The most considerable "failure" is observed in relation to such effects as the increase in the firms' competitiveness, the growth the domestic market share and the increase of investment in new equipment. Negative results concerning the last indicator seem quite surprising to us as a large set of tax incentives in Russia are principally intended to stimulate firms' investment activity. At the same time in contrast to a number of foreign studies, which examined a significant impact of tax incentives on input characteristics of innovation activity, first of all R&D expenses (see, e.g., Lokshin, Mohnen, 2012; Bodas Freitas et al., 2015), in Russia we can see no tangible input additionality of such measures. Slightly noticeable additionality effect of tax instruments relate to scale and scope additionality (the growth of investment in ongoing projects and the increase of the acceptable payback period). Note that the positive impact of tax support on scale and scope of ongoing projects in contrast to initiation of the new ones rather often was identified in economic literature (Guellec, Van Pottelsberghe, 2003; Jaumotte, Pain, 2005; Simachev et al., 2014b).

A detected clear dominance of financial instruments over tax incentives in most additionality effects, in our opinion, should not be considered as an exhaustive evidence of the inefficiency of tax measures and even more as a convincing argument in favor of abandonment of this element of the innovation policy. Indeed, the set of tax instruments obtains a number of important advantages.. Actually, they are potentially available for a wider range of recipients than direct funding instruments, other things being equal they are associated with lower implementation and administration costs (Simachev et al., 2014b), do not involve government intervention in market mechanisms and, what is important, are not directly linked to the budget allocation process (Gokhberg et al., 2014). It is also important that tax measures and public funding instruments have substantially different beneficiaries. Finally, tax incentives in a noticeably less degree produce a crowding out effect (replacement of private funds by public ones - see, e.g., David et al., 2000; Jaumotte, Pain, 2005) which is confirmed by the results of our study.

REFERENCES

Aschhoff, B. (2010): "Who gets the money? The dynamics of R&D project subsidies in Germany," Journal of Economics and Statistics (Jahrbücher für Nationalökonomie und Statistik), vol. 230(5), pp. 522–546.

Baghana, R. (2010): "Public R&D Subsidies and Productivity: Evidence from firm level data in Quebec," United Nations University - Maastricht Economic and social Research and training centre on Innovation and Technology, working paper, August 2010.

Bodas Freitas, I., F. Castellacci, R. Fontana, F. Malerba and A Vezzulli. (2015). "The additionality effects of R&D tax credits across sectors: A cross-country microeconometric analysis". Centre for Technology, Innovation and Culture, University of Oslo, working paper, March 2015

Busom, I and A.Fernandez Ribas. (2008). "The impact of firm participation in R&D programmes on R&D partnerships," Research Policy, vol. 37(2), pp. 240–257.

Cantner, U. and S. Kösters. (2015). "Public R&D support for newly founded firms – effects on patent activity and employment growth," Journal of Innovation Economics & Management, vol. 1(16), pp. 7-37.

David, P., B. Hall and A. Toole. (2000). "Is Public R&D a Compliment or a Substitute for Private R&D? A Review of Econometric Evidence," Research Policy, vol. 29 (4), pp. 497-529.

Falk, R. (2006): "Behavioural additionality in Austria's industrial research promotion fund (FFF)," Chapter 3 in: Government R&D Funding and Company Behaviour: Measuring behavioural additionality, Paris, OECD, pp. 59-74.

Falk, R. (2007): "Measuring the effects of public support schemes on Firms innovation activities," Research Policy, vol. 36(5), pp. 665–679.

Fier, A. and O. Heneric. (2005). "Public R&D Policy: The Right Turns of the Wrong Screw? The Case of the German Biotechnology Industry", ZEW Discussion Papers, working paper, Đ 05-60.

Fier, A., B. Aschhoff and H. Löhlein. (2006). "Detecting BehaviouralAdditionality: An Empirical Study on the Impact of Public R&D Funding on Firms' Cooperative Behaviour in Germany," ZEW Discussion Papers, working paper, Đ 06-037.

Georghiou, L., K. Malik and H. Cameron. (2005). "DTI Exploratory study on behavioural additionality," PREST, Manchester Business School and University of Manchester, working paper.

Gokhberg, L., G. Kitova and V. Roud. (2014). "Tax Incentives for R&D and Innovation: Demand versus Effects," Foresight-Russia, vol. 8(3), pp. 18-41.

Guellec D. and B. Van Pottlesberghe. (2003). "The impact of public R&D expenditure on business R&D," Economics of Innovation and New Technologies, vol. 12(3), pp. 225-244.

Higher School of Economics, Interdepartmental Analytical Center. (2009): "Evaluation of Anti-crisis Measures Supporting the Real Sector of the Russian Economy," Voprosy Ekonomiki, vol. 5, pp. 21-46.

Idea Consult. (2009): "Does Europe change R&D-behaviour ?Assessing the behavioural additionality of the Sixth Framework Programme", retrieved on 06 May 2016 from https://ec.europa.eu/research/evaluations/pdf/archive/fp6-evidence-base/evaluation_studies_and_reports/evaluation_studies_and_reports_2009/assessing_the_behavioural_additionality_of_the_sixth_framework_programme.pdf

Jaumotte, F. and N. Pain. (2005). "An Overview of Public Policies to Support Innovation," OECD Economic Department, working paper, Đ 456.

Kuzyk, M. and Yu.Simachev. (2013). "Russia's Innovation Promotion Policies: Their Evolution, Achievements, Problems and Lessons," Chapter 6.4 in: Russian Economy in 2012. Trends and Outlooks (Issue 34), Đ.: Gaidar Institute Publishers, pp. 435-479.

Lohmann, F. (2014). "The Additionality Effects of Government Subsidies on R&D and Innovation Activities in the Aviation Industry. A Project Level Analysis", retrieved on 06 May 2016 from http://essay.utwente.nl/64836/1/Lohmann_MA_MB.pdf

Lokshin, B. and P. Mohnen. (2012). "How effective are level-based R&D tax credits? Evidence from the Netherlands,"Applied Economics, vol. 44(12). pp. 1527-1538.

Loof, H. and A. Heshmati. (2005). "The Impact of Public Funding on Private R&D Investment: New Evidence from a Firm Level Innovation Study," MTT Discussion papers, working paper, Đ 3.2005.

Marzucchi, A. and S. Montresor. (2013). "The Multi-Dimensional Additionality of Innovation Policies: A Multi-Level Application to Italy and Spain," SPRU Working Paper Series, working paper, Đ 2013-04.

Mau, V. (2010). "Economic Policy in 2009: Between the Crisis and Modernization," Voprosy Ekonomiki, vol. 2, pp. 4-25.
OECD. (2015). OECD Science, Technology and Industry Scoreboard 2015: Innovation for growth and society, OECD Publishing, Paris.

Simachev, Yu., M. Kuzyk and V. Feygina. (2014a). "Fostering firms' innovation activities in Russia. Empirical analysis of beneficiaries and effects," in Yasin, Ev.G (Ed).: XIV April International Academic Conference on Economic and Social Development, vol. 3, pp. 369-388.

Simachev, Yu., M. Kuzyk and V. Feygina. (2014b). "Public Support for Innovation in Russia: What Can We Say about Tax Incentives and Public Funding?" Russian Management Journal, vol. 12(1), pp. 7-38.

Simachev, Yu., M. Kuzyk and V. Feygina. (2014c). "R&D Cooperation between Russian Firms and Research Organizations: Is There a Need for State Assistance?" Voprosy Ekonomiki, vol. 5, pp. 4-34.

Simachev, Yu., M. Kuzyk and D. Ivanov. (2012). "Russian Financial Development Institutions: Are We on the Right Way?" Voprosy Ekonomiki, vol. 7, pp. 4-29.

Wanzenbock, I., T. Scherngell and M. Fischer. (2013). "How do firm characteristics affect behavioural additionalities of public R&D subsidies?" Technovation, vol. 33 (2-3), pp. 66-77.

AUTHORS

YURI SIMACHEV
Russian Science Foundation
Moscow, Russia, 14/3, Solyanka str.,

MIKHAIL KUZYK
Interdepartmental Analytical Center
Moscow, Russia, 31/29, Povarskaya St., POB 35

NIKOLAY ZUDIN
Corresponding author
Interdepartmental Analytical Center
Moscow, Russia, 31/29, Povarskaya St., POB 35)

tel.: +79165715843
N_zud@mail.ru

HOW STRUCTURED DOCTORAL PROGRAMMES CHANGE THE LANDSCAPE OF DOCTORAL EDUCATION? AN EVALUATION APPROACH

JENS AMBRASAT AND JAKOB TESCH

BACKGROUND

Since the 1980s the system of doctoral training is in transition in Europe as well as worldwide (Nerad and Heggelund 2008). Following several criticism on an inefficient and often intransparent doctoral training, and long lasting completion times (e.g. WR 1996; Kehm 2008; Bartelse and Huismann 2008) an ongoing reform process is changing the organization, structure and purpose of doctoral training.

Within Europe the driving force of reforms stems from the Bologna Process in which doctoral education has been subsumed as third cycle of higher education (Berlin communiqué, 2003). This stream of reforms blends with policy desires on the employability of graduates for labor market demands within Europe's knowledge society (Mulvany et al. 2014). Structured doctoral programs are in the center of discussion about instruments operating in booth ways: to meet the goals of Bologna reform and labor market demands of the European knowledge society (EUA 2007; EU Commission 2011). The hopes linked to structured doctoral programmes (SDPs) are on restructuring doctoral training on the institutional level and thus improving quality and efficiency of doctoral training.

An increasing number of SDPs have been implemented at the institutional level of universities, faculties or research institutes and they were expected to change the structure of doctoral training significantly (Sursock and Smidt 2010). However, the actual change and the impact on the landscape of doctoral training were only scarcely examined. This might be primarily due to a lack of appropriate data for investigating and evaluating these changes in a comparative perspective. But, as we would argue, there is also a missing conceptual framework for such comparative analyses.

In theory most scholars contrast the emerging SDPs with rather idealized types of doctoral candidatures, either the master-apprentice model (Berning and Falk 2005; Janson et al. 2007; Kehm 2008; Hornbostel 2009) or the so-called individual doctorate (Wintermantel 2010). While the master-apprenticeship model points to the strong dependencies in the relationship between candidate and supervisor the term individual doctorate suggests that the candidate has one supervisor and no doctoral colleagues. Both ideal types may be a useful rhetoric antithesis to SDPs as they blaze the trail for political goals, but they do not constitute a suitable frame for empirical comparison and evaluation of SDP. So called individual doctorates or the master-apprenticeship model do not adequately describe the more complex landscape of doctoral training. In fact, there exist traditional established pathways of doctoral candidature, namely research assistants, scholarship holders, and external candidates, who share to different degrees the characteristics of an individual doctorate, a master-apprentice relation, or a structured doctoral training as it is aimed at by doctoral programs. These established pathways correspond to status groups and mirror form and source of financing as well as the degree of embeddedness within the scientific community. Thus candidates of these status groups differ not only in access to resources useful to successfully complete the PhD, but also in their prerequisites for an academic career (Laudel and Gläser 2008).

Although the distribution of the status groups depend on the structure of doctoral training within each country and therefore various across Europe, the named fundamental groups were discussed and can be found not only in Germany but in many countries across Europe (Ates et al. 2011; Huisman et al. 2002; Auriol 2010; Kehm 2007).

RESEARCH QUESTION

Surprisingly, these features of doctoral status groups have never been conceptualized systematically and thus influences on doctoral training have been ignored so far.

Within this paper we suggest a conceptual framework for comparing the different contexts of doctoral training including traditional pathways and structured doctoral programmes. In particular, we argue the change invoked by emerging SDPs can only be evaluated appropriately in the interplay with traditional pathways.

Thus, our research question can be refined as follows: How emerging SDPs change the established structure of doctoral training for the backdrop of the traditional pathways? In general, SDPs are expected to improve the quality of supervision and provide institutional settings that promote an efficient and successful doctoral training, e.g. by course offers, recorded agreements, timelines, etc. In more detail, it can be expected that the effects of SDPs differ between traditional paths of doctoral training.

The prototype of a SDP candidate is a member who holds a scholarship for mostly three years that gives her the opportunity to fully participate in the program including all courses, meetings and other support. Since scholars have the biggest time budget it can be expected that supporting structures of doctoral programs can unfold their full effects.

The situation is something different for research assistants who work in a third-party funded research project or at the department with duties in education and administration and thus have more limited time resources. For research assistants the situation as SDP member entails both risk and chances. On one hand it might let them off the close and sometimes intransparent (master-apprentice) relationship with the professor, which might improve the quality of supervision not at least by contacting a second or even more supervisors. On the other hand, the curriculum and other duties resulting from a SDP membership constitute an extra load in terms of time resources.

Finally, the situation of external candidates, who are usually not very well embedded in academic communication and networks, bears also substantial potential for improvement. External candidates should more than all others profit from an integration in a SDP. In particular, it is expected that their situation of supervision improves, the degree of integration increases and curricular courses will be made accessible.

In addition to the question of whether SDPs fulfil the expected goals, we are interested in how stable the established structure of traditional pathways is. Does it continuously shape the landscape and determine the structure of doctoral training or is this former structure being overridden by the new SDP structure?

DATA AND METHOD

To substantiate our argument we use data from a large longitudinal study with about 9.,000 doctoral candidates who took part in our panel study ProFile between 2009 and 2015. The panel captures doctoral candidates who were enrolled at one of 14 cooperation institutions in Germany. Our key explanatory variable is the membership in a structured doctoral program. SDP-members have been oversampled to enable better comparisons with non-members who are dominant in the entire population. Moreover, we distinguish research assistants, scholarship holders and external candidates with and without job as traditional status groups.

As dependent variables we examine several aspects of doctoral training that point to the formalization and supporting structure of the institutional environment. In particular, we look at recorded agreements, multiple supervisor constellations, exchange intensity, course attendances, and the time candidates invest into work on their thesis. To demonstrate differences between traditional pathways of doctoral training and between program members and non-members, we apply multiple regression models where we control for subject field and year of survey participation.

RESULTS

We evaluate the role of SDPs in structuring and formalizing the doctoral training in the interplay and in comparison with traditional pathways.

Results confirm our general argument that changes in doctoral training invoked by emerging SDPs can only be adequately evaluated when recognizing traditional pathways of doctoral training that shaped the landscape so far. Regardless of any SDP membership, status groups already highly determine the structural context of doctoral training. Structural elements like recorded agreements and multiple supervisors are determined mainly by status groups. Likewise, course attendance and the assessed quality of supervision eminently depend on the contexts formed by traditional pathways.

Concerning the impact of SDPs, we can show that they meet almost all expectations in reshaping doctoral education. They significantly improve the formal and structural environment of doctoral training and lead indeed to an expanded course offer of scientific, interdisciplinary and transferable skills courses.

However, the impact of SDPs differs across status groups that means, SDPs develop their intended effects on supervision in dependence on the traditionally evolved structure of doctoral training. Most of all research assistants profit from an SDP environment and become disconnected from a close master-apprentice relationship when entering SDP contexts. Significantly and a little bit surprising, they are able to increase the time they spend for working on the thesis.

In total, the following picture emerges. Structured doctoral programs meet a well-established background structure of traditional pathways essentially determined by status groups. Thereby they not really create a new infrastructure but develop and improve existing structures, which also positively affects non-SDP members as well.

Our analyses lead to a better understanding of the different pathways of doctoral candidature shedding light on the diverse situation where structured doctoral programmes function in the interplay with traditional pathways of doctoral education. Our proposed framework for comparisons of different contexts of doctoral training allows estimations of future developments too. If the proportion of SDPs continues to increase, further improvements of doctoral training are most probable.

REFERENCES

Ates, Gülay et al. 2011. Eurodoc 2010. The first Eurodoc Survey on Doctoral Candidates in twelve European Countries . Brüssel: Eurodoc.

Auriol,L. (2010), "Careers of Doctorate Holders: Employment and Mobility Patterns", OECD Science, Technology and Industry. Working Papers, No. 2010/04, OECD Publishing, Paris.

Berlin Communiqué (2002) Realising the European Higher Education Area. Berlin: Berlin Communiqué.

Berning, Ewald and Susanne Falk. 2005. Abschied vom "Meister-Schüler-Verhältnis"? Strukturwandel in der Promotionsbetreuung. Forschung & Lehre 12: 686-687.

European University Association (2005), Doctoral Programmes for the European Knowledge Society: report on the EUA doctoral programmes project, Brussels, EUA. Available at: http://www.eua.be/eua/jsp/en/upload/Doctoral_Programmes_Project_Report.1129278878120.pdf

EUA. 2010. Salzburg II Recommendations. Available at:
http://www.eua.be/Libraries/publications-homepage-list/Salzburg_II_Recommendations

EU Commission. 2011. Principles for Innovative Doctoral Training. Available at: http://ec.europa.eu/euraxess/pdf/research_policies/SGHRM_IDTP_Report_Final.pdf research_policies/Principles_for_Innovative_Doctoral_Training.pdf. Accessed 31st May 2013.

Guth, Jessica. 2006. The Bologna Process: The Impact of Higher Education Reform on the Structure and Organisation of Doctoral Programmes in Germany. Higher Education in Europe. 31(3): 327-338.

Hochschulrektorenkonferenz
Hornbostel, Stefan. 2009. Promotion im Umbruch - Bologna ante Portas. In: Martin Held, Gisela Kubon-Gilke and Richard Sturn (Eds.): Jahrbuch Normative und institutionelle Grundfragen der Ökonomik, Band 8, Bildungsökonomie in der Wissensgesellschaft.

Huisman, Jeroen; Weert, Egbert de; Bartelse, Jeroen. 2002. Academic Careers from a European Perspective: The Declining Desirability of the Faculty Position. The Journal of Higher Education 73(1): 141-160.

Janson, Kerstin; Schomburg, Harald; Teichler, Ulrich. 2007. Wege zur Professur. Qualifizierung und Beschäftigung an Hochschulen in Deutschland und den USA. Münster (Westf.): Waxmann.

Kehm, Barabara. 2007. Quo Vadis Doctoral Education? New European Approaches in the Context of Global Changes. European Journal of Education 42(3): 307-319.

Kehm, Barbara. 2008. Germany. In: Nerad and Hegelund (Eds): Toward a Global PhD. Forces & Forms in Doctoral Education Worldwide. University of Washington Press.

Laudel, Grit, and Jochen Gläser, 2008. From apprentice to colleague: the metamorphosis of Early Career Researchers. Higher Education 55 (3): 387-406.

Mulvany MJ, Jonsson R, Lackovic Z. 2014. European biomedical and health related PhD education: A tale of two cultures. In: Proceedings from International Conference on Developments in Doctoral Education & Training pp. 43-53.

Nerad and Hegelund (Eds) 2008.Toward a Global PhD. Forces & Forms in Doctoral Education Worldwide. Seattle (WA): University of Washington Press.

Sursock, Andrée and Hanne Smidt. 2010. Trends 2010: A decade of change in European Higher Education. EUA
Wintermantel, Margret. 2013. Promovieren heute. Zur Entwicklung der deutschen Doktorandenausbildung im europäischen Hochschulraum. Hamburg: Edition Körber.

Wissenschaftsrat. 1996. Empfehlungen und Stellungnahmen 1995. Köln: Wissenschaftsrat.

AUTHORS

JENS AMBRASAT
DZHW - German Centre for Higher Education Research and Science Studies, Berlin
Deutsches Zentrum für Hochschul- und Wissenschaftsforschung GmbH
Standort Berlin
Schützenstraße 6a
10117 Berlin
Gemany

Phone: +49 30 2064177-0
e-mail: ambrasat@dzhw.eu

JAKOB TESCH
DZHW - German Centre for Higher Education Research and Science Studies, Berlin
Deutsches Zentrum für Hochschul- und Wissenschaftsforschung GmbH
Standort Berlin
Schützenstraße 6a
10117 Berlin
Gemany

MEASURING AND UNDERSTANDING INTERDISCIPLINARITY IN COMPUTER SCIENCE DOCTORAL PROGRAMS

MARÍA DEL CARMEN CALATRAVA MORENO AND HANNES WERTHNER

Interdisciplinarity has become a major topic in academic and policy oriented discussion on knowledge production and research funding, and the reason for major policy reforms. European research funding institutions are placing a clear emphasis on interdisciplinar- ity as a means to spur innovation (e.g., the Research and Innovation program Horizon 2020), thereby strengthening the competitiveness of the European Research Area. Euro- pean higher education institutions have responded to the need of educating future interdis- ciplinary scientists by developing new forms of doctoral education — doctoral schools and colleges as alternatives to the traditional PhD program — to prepare interdisciplinary early career researchers.

As a result, there is a demand for both the monitoring and the understanding of progress towards greater interdisciplinary research. This requires, on the one hand, the definition and development of criteria and tools for the measurement and quantification of interdisciplinary research, and on the other hand analyses of factors and facilitators of interdisciplinary research.

We conducted a study that includes both perspectives and focuses on doctoral educa- tion. It was carried out in three Austrian doctoral programs that run parallel in the same institution but have different curricular approaches: a traditional European doctorate and two structured doctoral programs–one which is multidisciplinary within computer science and another co-organized by three faculties. We focus on the field of computer science, which has become a field of its own in spite of its interdisciplinary origins (Tedre, 2014). Although it is a highly collaborative field because of its multiple applications, discipline- specific research is the most common kind of research in many areas of computer science.

QUANTIFYING INTERDISCIPLINARY RESEARCH

Since interdisciplinary research is often conceptualized as the integration of diverse knowl- edge, one of the most common methods for its measurement is citation analysis, in which an exchange or integration among fields is captured via discipline-specific citations refer- ring to other fields. Such an approach is especially useful for large-scale measurements, and it is usually used in combination with a predefined taxonomy of disciplines to classify publications into disciplinary fields (Leydesdorff, Carley, & Rafols, 2013; Porter & Rafols, 2009; Rafols, Leydesdorff, OHare, Nightingale, & Stirling, 2012). Unfortunately, the use of such taxonomies presents two additional obstacles to obtaining accurate results. First, the taxonomy of disciplines needs to accurately describe the research landscape (e.g., with the right level of detail of clusters of disciplines, updated to include emerging disciplines). Al- though there is no consensus as to which is the best taxonomy (National Research Council, 2010; Rafols & Leydesdorff, 2009), the one utilized by Web of Science is the one most widely used (Bensman & Leydesdorff, 2009; Pudovkin & Garfield, 2002). Second, each citation needs to be categorized into at least one discipline within the taxonomy. If citations remain uncategorized, they will not be taken into account in the analysis. The more citations that remain uncategorized, the less accurate the IDR measurement will be.

Although the major consequences of missing data in bibliographic datasets have been acknowledged in the literature (Moed, Burger, Frankfort, & Van Raan, 1985), the exhaus- tive categorization of all references within a dataset into disciplinary fields remains an open issue that is under-discussed. In order to tackle this problem, we propose a method which acknowledges missing data and determines the associated uncertainties (Calatrava Moreno, Auzinger, & Werthner, 2016). Our method is an extension of the Rao-Stirling diversity index (Porter & Rafols, 2009), which not only captures the variety and balance of the disciplines cited by a paper, but also their disparity using a measure of similarity between disciplines. To capture the effects of missing data, we compute the range in which the Rao- Stirling index can vary when the uncategorized references are assigned to relevant arbitrary disciplines. In other words, this extension of the Rao-Stirling index encodes the uncertainty caused by missing data as an interval. The main benefit of this uncertainty interval is that it acts as a confidence indicator of the results delivered by the Rao-Stirling index. On the one hand, publications with a low proportion of uncategorized references have correspond- ingly small uncertainty intervals, implying a more reliable measurement of the Rao-Stirling index. On the other hand, publications with a high proportion of uncategorized references have correspondingly large uncertainty intervals, indicating an unreliable measurement of the Rao-Stirling index. This uncertainty interval would allow interdisciplinary research an- alysts to assess the validity of their bibliographic data and discard publications with high uncertainty from the analysis. Our contribution is a first approach to measure interdisci- plinary research, taking into account the incompleteness of bibliographic data.

We used the Rao-Stirling index and our uncertainty interval to measure the inter- disciplinarity of the publications of 195 students from the three doctoral programs. The interdisciplinarity of the three programs was calculated by aggregating the results of their respective students.

ANALYSIS OF DOCTORAL STUDENTS' EXPERIENCES

Using the results of the Rao-Stirling index and our uncertainty interval, the 15 most inter- disciplinary students of the three doctoral programs were selected to participate in in-depth semi-structured interviews that helped us to understand how dispositions and experiences of students and factors of the different programs affect the circumstances and process of becoming an interdisciplinary early career researcher. Moreover, we quantitatively analyzed the distribution of their references to different disciplines in the sections of their publica- tions, and assessed the function and importance of other areas of knowledge in the research of interdisciplinary students.

The data indicate that besides interdisciplinary doctoral structures, other factors such as student values, motivations, as well as previous skills and knowledge, interacted with policies and program structures including type of funding and supervisor expectations to play an important role in interdisciplinarity at the doctoral level. The factors and processes identified in our analysis not only play an important role in giving rise to interdisciplinary research — even in programs without interdisciplinary focus — but also compromising the interdisciplinary goals of interdisciplinary programs. Moreover, among the highly interdis- ciplinary students, there were substantial differences in their individual characteristics and experiences in their doctoral program. We identified three patterns of doctoral interdisci- plinarity in computer science: integrative, disciplinary, and specialist.

The contributions of this study on doctoral experiences extend the literature of inter- disciplinary education and add to the complexity of the two existing models of interdis- ciplinary education: one in which individuals are trained in a specific discipline and later engage in interdisciplinarity, and another that assumes that individuals are already trained as interdisciplinary researchers (Holley, 2006; Klein, 1990; Messmer, 1978). We offer a third variant and relate the different models to both the individual and the doctoral processes. Moreover, our contributions derive from both bibliographic data and students' descriptions, opening a new way to further research and policy initiatives to facilitate the development of interdisciplinary early career scientists.

REFERENCES

Bensman, S. J., & Leydesdorff, L. (2009). Definition and identification of journals as bibliographic and subject entities: Librarianship versus ISI Journal Citation Reports methods and their effect on citation measures. Journal of the American Society for Information Science and Technology, 60(6), 1097–1117. doi: 10.1002/asi.21020

Calatrava Moreno, M. C., Auzinger, T., & Werthner, H. (2016). On the uncertainty of in- terdisciplinarity measurements due to incomplete bibliographic data. Scientometrics, 107(1), 213–232. doi: 10.1007/s11192-016-1842-4

Holley, K. A. (2006). The cultural construction of interdisciplinarity: Doctoral student socialization in an interdisciplinary neuroscience program. University of Southern California. (Unpublished doctoral dissertation)

AUTHORS

MARÍA DEL CARMEN CALATRAVA MORENO
E-Commerce Group,
Institute of Software Technology and Interactive Systems,
TU Wien, Vienna
Favoritenstrasse 9-11, 1040 Vienna
Tel.: +43 (1) 588 01 -188307
Email: calatrava@ec.tuwien.ac.at
Austria.

HANNES WERTHNER
E-Commerce Group,
Institute of Software Technology and Interactive Systems,
TU Wien, Vienna
Austria.

THE EFFECTS OF GERMAN RESEARCH FOUNDATION GRANTS ON CAREER PATTERNS

RICHARD HEIDLER

INTRODUCTION

The contribution presents the results of an analysis and comparison of the effects of five types of grants from the German Research Foundation on the careers patterns of scientists. The study uses an innovative research design that combines a cohort design comparing granted with rejected applications and a CV-analysis. Additionally, methodological challenges and benefits resulting out of such an approach are discussed.

A still growing share of funding in the German research system is allocated by competitively awarded research grants. This process is accompanied by two conflicting trends: (1) a stagnation in the number of full professors with permanent contracts at German universities and (2) an expansion of the amount of young investigators. The likelihood to gain a professorship is under challenge because of a larger group of people competing for the highest levels of the academic career ladder. Similar trends of an increasing share of untenured Postdocs with an unclear career prospective are observable in other countries, too (ESF 2015, Cyranosky et al 2011). Thus, the impact of research grants as a factor for shaping and promoting research careers is of growing importance for the individual researcher.

Research funders are affected by these changing expectations, too. For a long time a crucial measure of funding success has been the question whether the funding decision succeeded to select the best proposals and promote projects. Recently the question is gaining weight to what extent the funding organizations manage to contribute to the long-term career placement of outstanding scientists with their programs.

RELATED STUDIES

These changes in the science system are mirrored in the dominant research designs of funding evaluation studies. While the effects of research grants on scientific productivity and impact are well studied, oftentimes by bibliometric methods (Langfeldt

et al 2015, Neufeld 2013), the implications of research grants for scientific career decisions and career performance are much less understood (Bloch et al. 2014). Only recently studies for different funding programs e.g. from the University of Vienna in Austria (Reimann & Wysocki 2015), the Danish research council (Bloch et al. 2014), the ERC (Huber, Wegner & Neufeld 2015) or from a set of European funding organizations (ESF 2015) track the long-term career success of granted applicants, sometimes comparing with the rejected ones. These studies assume that the micro-behavior and lasting success of scientists is shaped by the specificities of funding arrangements they apply for and by the subsequent funding decision. This career effect can occur by generating competencies and by pushing productivity through an expansion of researcher's resources, by network building but also by increasing the visibility and reputation of scientists. Some studies of that type evaluate the career effect of these funding schemes by comparing accepted and declined applicants, while controlling for confounding covariates with statistical models.

METHODOLOGICAL CHALLENGES

However, the validity and robustness of the results can be limited by two methodological challenges. Firstly, low response rates can occur when questionnaires are used to asses careers, which can produce a systematic dropout for non-successful applicants. A second challenge to the validity of these studies are diverging definitions and considerations of the scientific career age of the populations under study. An arbitrary choice of cohorts is problematic, since the career success is at least partly a direct function of the years in the science system since the PhD and of long-term trends of an improving or worsening career prospective.

To address the first problem, the proposed study proposes building on the CV-method to gather data instead of using questionnaires (Cañibano & Bozemann 2009). A comparable study following the described design based on secondary data is the above-mentioned study about the Danish science system, comparing rejected and granted applicants. The study builds on a statistical model that shows that the probability of obtaining a full professorship almost doubles for the successful applicants (Bloch et al. 2014).

ANALYSIS AND RESULTS

The presented study extends the approach from the Danish study to the German research system and the funding programs from the German Research Foundation (DFG) relevant for postdoctoral researchers. Results for five person-oriented funding schemes will be presented, which are tailored at or can be used for conducting research projects and pursuing careers. The studied programs differ in their funding goals, funding rate, grant volume, duration, application requirements and other aspects. They address different career stages and enjoy varying popularity in different scientific disciplines. The funding programs which will be studied and compared in detail are the "research fellowship", the "temporary position" ("Eigene Stelle"), the "Emmy Noether-program", the "Heisenberg fellowship" and the "Heisenberg professorship". Within these programs, granted and rejected applicants are contrasted. This will be done by using the CV-method to gather standardized information from 1,133 curriculum vitae. To this purpose, CVs attached to later applications are used and additional (or more recent) CVs from public sources and websites are included. For the cohorts from the years 2007/2008 the past and future careers are tracked by a standardized coding of

information on the PhD, research stage, research position, habilitation, junior professorship, occupational sectors and home country.

Based on this data, an econometric analysis models the treatment effect of the five funding programs on future career stages and on the chances to become a full professor. In addition, for selected programs the chances that the applicant will later continue his or her career in a foreign country or leave the science system will be modeled.

A supplementary in-depth analysis substantiates the results and addresses the question of the dependency of the career progress upon the "scientific age" and the cohort chosen for the study. Therefore, an automated long-term career tracking is presented, using current researcher titles and sliding cohorts from the DFG-databases to corroborate the results from an extended perspective. In a first step, the validity of title analysis for assessing the career stage is tested, based on a comparison of the database titles and the hand-coded CV-data for the cohort of 2007/2008. Thereafter year wise cohorts of rejected and granted applicants from 2001 to 2015 are tracked and compared based on their actual title. The results will be triangulated with the CV-data analysis.

DISCUSSION

The results ground on a research design modelling the effect of the funding decision for defined cohorts of granted and rejected applicants within a multivariate statistical model, based on exhaustive secondary data. This leads to a methodologically confirmed analysis avoiding some of the pitfalls occurring in studies of the career effects of funding programs.

Bloch, C., E. K. Graversen and H. S. Pedersen (2014): "Competitive Research Grants and Their Impact on Career Performance," Minerva, vol. 52, pp. 77-96.

Canibano, C. and B. Bozeman (2009): "Curriculum vitae method in science policy and research evaluation: the state of the art," Research Evaluation, vol. 18, pp. 86-94.

Cyranoski, D., N. Gilbert, H. Ledford, A. N. Nayar and M. Yahia (2011) Education: "The PhD factory," Nature News Feature vol. 472, pp. 276-279

European Science Foundation (2015): Career Tracking of Doctorate Holders Pilot Project Report. Strasbourg: Ireg.

Huber, N., A. Wegner and J. Neufeld (2015): "MERCI (Monitoring European Research Council's Implementation of Excellence): Evaluation Report on the Impact of the ERC Starting Grant Programme," iFQ Working Paper, No.16. Berlin: iFQ.

Langfeldt, L., C. Bloch and S. G. Walter (2015): "Options and limitations in measuring the impact of research grants - evidence from Denmark and Norway," Research Evaluation, vol. 24, pp. 256-270.

Reimann, R. and E. L. Wysocki (2015): „Karrierewege von FWF-ProjektleiterInnen an der Universität Wien," in Ash, M. G. and J. Ehmer (Ed.): Universität – Politik – Gesellschaft, Vienna: Vienna University Press 2015.

AUTHOR

RICHARD HEIDLER
German Research Foundation
Kennedyallee 40
53175 Bonn
Tel. +49 (228) 885-2925
e-Mail: richard.heidler@dfg.de

THE EVALUATION OF THE AUSTRIAN START PROGRAMME : AN EXAMPLE FOR A SUCCESSFUL SOPHISTICATED MULTI-METHOD APPROACH

SARAH SEUS AND EVA HECKL

INTRODUCTION

The START Programme is one the most prestigious research grant for post-doctoral researchers in Austria. It targets excellent young researchers of any disciplines, who do not yet hold a permanent professorship. The objectives of the START programme are twofold: „researchers should be given the long-term and extensive financial security to plan their research and to build up or consolidate their own research groups thereby qualifying themselves for senior research positions", especially as university professors within Austria or abroad (Austrian Science Fund: https://www.fwf.ac.at/de/forschungsfoerderung/fwf-programme/start-programm/). The START Programme was created in 1996 and is managed by the FWF (Austrian Science Fund). It is one of Austria's most generously supported research programme with the provision of a grant of up to 1.2 million Euros for a time up to six years. This grant can be flexibly spent on personnel of the research group and equipment. Between 1996 and 2014, a total of 114 START grantees were funded. The sole selection criteria being scientific excellence, the grant has been allocated to researchers from the overall research discipline spectrum, covering natural sciences, life sciences, social sciences as well as humanities.

The evaluation of the START programme was commissioned by the FWF and carried out between October 2014 and April 2016 by the Fraunhofer Institute for System and Innovation Research and the Austrian Institute for SME Research (KMU Forschung Austria).[1] The objectives of the evaluation were to assess the achievements of the programme, especially on output and outcome level. Furthermore it should provide information on the role the programme plays in the FWF's overall funding portfolio for post-doctoral research.

The aim of this contribution is to present the methodological design that combines elements of quantitative impact analysis with qualitative elements stemming from different sources. On the basis of selected key findings of the evaluation the strengths of this approach is discussed. In particular, it shows the role of triangulation (e.g. Leeuw 2009 et al.) and its potential to either reinforce or discard preliminary conclusions and provide further explanatory variables for the measurements.

THE STUDY DESIGN

As the evaluation was intended as an impact evaluation, the study design should on the one hand quantify the results as far as possible and on the other hand give indication whether those documented effects could be attributed to the START funding. In consequence, the evaluation design chosen was a quasi-experimental one, which incorporated a counterfactual in form of a control group. The strength of an experimental design is its potential to assess whether measured effects can be attribution of programme. However, such a design does not provide explanation how and why policy interventions work (or do not work) (Mayne 2012, 2001; Morton 2015). The opening of this black box is however crucial for an in depth understanding of the programme, and also to provide meaningful policy recommendations. In order to counterbalance the weaknesses of traditional quantitative impact analysis (Leeuw 2009 et al.), the evaluation was complemented by the analysis of further data, using different data collection methods and sources (mixed method approach). With the aim of triangulating the evidence, the analysis combined different data sources as well as qualitative and quantitative analytical methods, wherever possible and deemed meaningful.

A control group was created at the beginning of the evaluation. It consisted of researchers who showed comparable characteristics (discipline, gender, scientific age, affiliation to Austrian research organisations, publication output and citation rate) to the START grantees at the time of the START award. For each START grantee "twins" were identified based on information in the bibliometric database Scopus.

The control group was then used for two different types of analysis: first, an analysis of selected bibliometric indicators was performed on the two groups and results compared. For this purpose only one twin, called the "main twin" was used. This provided information on the scientific output of the programme participants compared to their non-funded counterparts. In order to also reflect the changes of performance over time, the analysis was furthermore divided into three periods of analysis: before, during and after the reception of the START grant. Secondly, an online survey was sent to both groups with the aim to quantify the effects of the START grant on the career aspects, especially the sector and

1 The study included the evaluation of the START programme and the Wittgenstein Award. In this paper only the evaluation approach for the START programme is discussed.

location of employment, the position hold and the pace of career development. (In order to grantee a sufficiently high success rate, up to three twins per START grantee were identified and received an invitation to participate in the online survey.) The online survey was sent to 114 START grantees and 307 control group participants; we received 94 responses of START grantees and 75 responses from the control group.

In order to get a deeper understanding of the causal mechanisms of the effects of the funding, the results of the control group analysis was complemented by further evidence of mostly qualitative, but also of quantitative nature: Eight case studies were made of START projects that combined the views of START grantees, representatives of host institutions of START projects, and START project group members (23 interviews in total).. Interviews with institutional stakeholders (FWF, Ministry) and analysis of programme documents, project reports, and monitoring data complemented the evidence base of the evaluation. Furthermore, an online survey was sent to a "comparison group" that consisted of unsuccessful applicants to the START Programme. These applicants came until the last selection stage, the "hearing" and therefore show very similar characteristics with regards to the scientific performance than the START grantees. (49 questionnaires sent; 25 responses).

new and unconventional research fields. Indeed 75% of the respondents to the survey to grantees stated that START was conducive to accessing new fields of research. From the interviews of START grantees we learnt that START projects often provided the room for the development of new methods. Hence, one reason for the downfall of citation rate might be that some START grantees have ventured a re-orientation of their research towards new or more unconventional research fields for which he/she is not (yet) visible in the research community. The slight rise of the citation rate in the period after the START funding can be seen as an indicator for a promising outlook with regard to future scientific performance of the former START grantees, once they have settled in their new research field. That the START programme is beneficial to venturing new research areas is supported by the self-assessment of the comparison group: 50% of the respondents state that they were not able to access new fields of research, as planned in the START project.

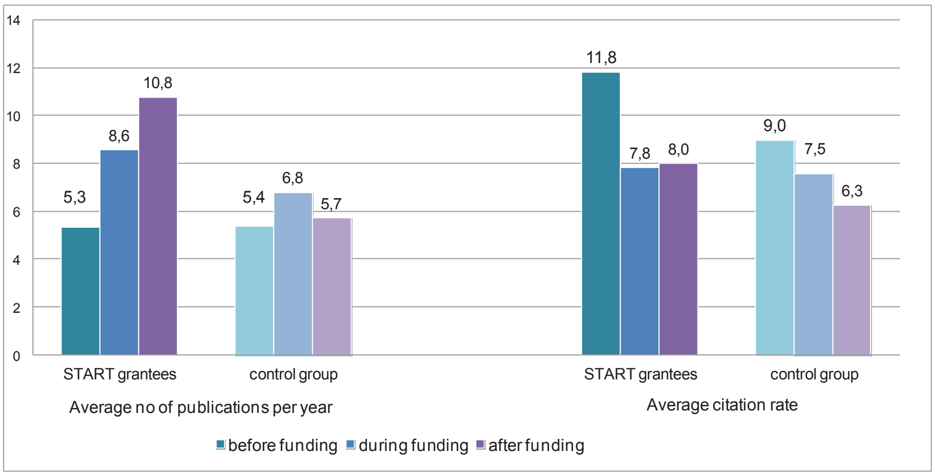


Figure 1: Publications and citation rate

EVALUATION RESULTS

The triangulation of different data collection methods and data sources has been especially fruitful, as it provides explanatory variables and interpretations guidelines, as the following examples show.

The evaluation was able to show that the scientific performance of the START grantees increased continuously for almost all measured indicators in the two periods after the reception of the grant. We measured the numbers of publications, the citation rate and three indicators measuring the co-publication activities (co-authors, co-organisations, co-countries) (see Fig.1 and 2). The only indicator with a harsh decrease from the period before the reception of award and the period during the START funding was the citation rate. The interviews and the survey answers of the START grantees provided a possible explanation for this phenomenon, namely that the START grant provided scope for testing

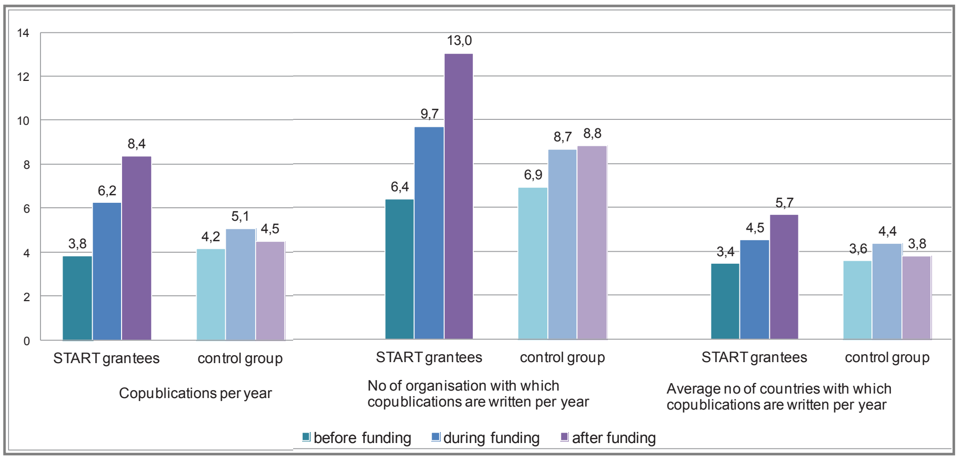


Figure 2: Co-publications, collaborating organizations and countries
Source: Seus, Heckl, Bührer 2016

One of the programme's objectives is the promotion of career advancement of the grantees and especially the ‚permanent integration' of the grand holders into the (Austrian) research system. Usually this means a full professorship position at a university. The overall majority of survey respondents and interviewees perceived the START grant as a motor for the career development and are convinced that START strengthened their career prospects and can also provide example for successful career effects. 60% of the respondents think that they would not have achieved their current position without START. The survey results confirm this assessment: All START grantees stay in the research system and almost two third of them stayed in an Austrian research institution. 80% of the START grantees hold or have held a professorship position (see fig. 3., all professorship positions are considered). However, for all for indicators, the control group shows a similar career pattern and does not differ significantly. Especially with regard to the pace towards a professorship, it has to be concluded from the survey results that the START Programme has no influence. START grantees do not get appointed to a professorship earlier than the control group or candidates for the START Programme. However, the interviews and the survey answers suggest that START grantee have strengthened their negotiation power and were able to get positions with better working conditions.

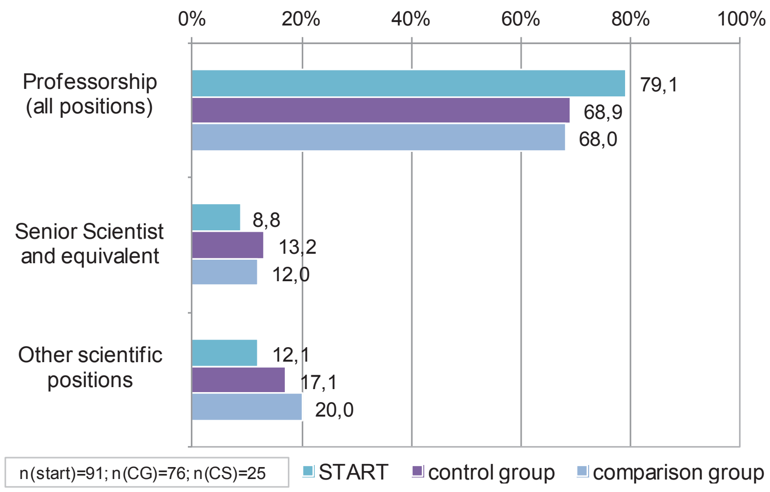


Figure 3: Highest position in the research system reached to date

With regard to the discussion whether the START applicants should be forced to submit at the same time an application for the Starting Grant of the European Research Council (ERC), the used approach was able to provide evidence. While there are strong voices in the Ministry and among the START grantees to abolish the obligation for a double application, the representative survey among START grantee provides another picture: the majority of START grantees are in favour of the current double application procedures and would have sent an application for both programmes even without the obligation. The qualitative answers from START grantees specify the reasons why a double application is favoured: the additional work created is acceptable and the FWF takes the ERC deadlines into account in the START application planning and is open for discussion in specific cases. The obligation of double application has strongly contributed to today's high success rate of Austrian researchers in ERC grants. There is furthermore evidence that the double application also pushes the internationalisation of humanities that are traditionally less used to apply for European programmes.

REFERENCES

Leeuw F., Vaessen, J., 2009): "Impact evaluations and development – NONIE guidance on impact evaluation", Network of Networks on Impact Evaluation, Worldbank. Retrieved on 29 April 2016 from: http://siteresources.worldbank.org/EXTOED/Resources/nonie_guidance.pdf

Mayne J, (2001): "Addressing attribution through contribution analysis: using performance measures sensibly," Canadian Journal of Program Evaluation 16(1), pp. 1–24.

Morton, S. (2015): "Progressing research impact assessment: A 'contributions' approach", Research Evaluation 24 (4), pp. 405-419.

Seus, Sarah / Heckl. Eva / Bührer, Susanne (2016). Evaluation of the START Programme and the Wittgenstein Award. Zenodo. Retrieved on 29 April 2016 from: 10.5281/zenodo.50610

AUTHORS

SARAH SEUS
Fraunhofer ISI
Breslauer Str. 48
76139 Karlsruhe
Germany
E-Mail: ses@isi.fraunhofer.de
Tel.: 0049-721-6809-344

EVA HECKL
KMU Forschung Austria
Gußhaustr. 8
1040 Wien
Austria
Tel.: 0043-1-5059761

THE EVALUATION OF THE ECONOMIC IMPACT OF RESEARCH INFRASTRUCTURES IN OPEN INNOVATION AND RESEARCH ENVIRONMENTS : THE EVARIO PROJECT

BACH LAURENT AND WOLFF SANDRINE

Since the early 2000s, the development and the coordination of large research infrastructures (RIs hereafter) have been increasingly recognized by the governments as well as by the the EU as an essential element of the Science and Technology policies. Considering the amount of resources invested in RIs, the evaluation of their socio-economic impacts is increasingly needed to support policy decisions. Then (at least partly) specific evaluation methods are required to cope with the specificities of this type of policy tool.

This paper results from the work carried out under the EvaRIO project conducted from 2011 to 2013 (EvaRIO : Evaluation of Research Infrastructures in Open innovation and research systems - CSA project under EC FP7/2007_2013, Grant Agreement n° 262281). The project aimed at developing an evaluation framework and a set of specific evaluation methods and tools well suited to RIs in the currently changing context towards an open innovation and research environment.

More specifically, the presentation highlights some of the main methodological specificities of the EvaRIO approach, including the architecture of effects, presents some examples of the implementation of evaluation tools, as well as comments on some promising extensions of the approach in the perspective of open research and innovation systems. It also describes the scope of relevance of the approach together with its complementarity with other approaches, suggesting that it could be considered as a link between different methodologies.

According to EU official documents (e.g. EC 2010), the term "Research Infrastructures" refers to facilities, resources or services of a unique nature that are needed by the scientific and technological communities to conduct basic or applied research in the whole range of scientific and technological fields. This definition covers major equipment (such as telescope or synchroton) or group(s) of related instruments; knowledge based-resources such as collections, archives, structured information or systems related to data management (such as biological archives, mouse depositeries or clinical research components including cohorts of patient, clinical specialist and trials management system); information and communication technology-based infrastructures such as grid computing, networks and communications. RI may be 'single-sited' physical facilities (a single resource at a specific location), 'distributed' (a network of distributed resources). They may provide remote access (e.g. where 'virtual' access to a core facility is provided electronically). They share the following features: large research capacity, trans-national relevance, requiring sizeable investment, access based on scientific excellence and, generally, high operating costs.

The evaluation of the economic or socio-economic impact of RIs is by far not a completely new domain in the field of impact evaluation, but it is probably not as well developed as the evaluation of more standard "public R&D programmes". Then classical and frequently cited surveys on evaluation in this field (see for instance Jordan & Ruegg (2007) or Georghiou et al. (2002), MioiR and Technopolis (2012) or Arnold (2012) for the specific case of EU programmes), as well as guidelines for impact assessment of public actions from OECD, EC, or other international organizations are useful but not fully adapted to this specific context.

In order to try to take into account the complexity and the variety of impacts, many attempts have been made to elaborate some classification and overall analytical framework, but only a few of them go beyond a simple list of items. For instance, distinctions between broad categories of impacts often includes Science, Technology & Innovation, Economic Impact, Social Impact, Environmental Impact and the like such as in the frameworks recently proposed by the partners of the RIFI project (Curaj et Pook (2011)) or in Zuijdam (2011), Péro (2009).

Different studies (in particular in the UK) were inspired by the seminal SPRU-type work on the impact of science (B. Martin, A. Salter etc.), from which the UK laid down the foundations of the evaluation system used for some of the UK RIs (especially from the STFC - Science and Technology Facilities Council, Dougan - McCallie 2014).

A more contextual approach to the general "mapping" of effects is provided by the well-known "logic model" (see for instance the scheme proposed by Technopolis (2010) in the case of RIs in bio-medical sciences).

Obviously, the overall Cost-Benefit Analysis framework is a candidate, especially given the size of the investment required for large facilities. But beyond some general recommandations such as those provided in EC (2014) or EIB (2013), it has only been recently specifically adapted to the case of RIs particularly in Clarke et al. (2013), Florio et al. (2016), and mainly in an ex ante perspective.

One of the most encompassing framework has probably been proposed by Simmonds et al. (2013), distinguishing different phases of development and operation of RIs with corresponding categories of effects, and relating them to a whole range of available evaluation methods.

As regards evaluation studies per se, only a few comprehensive ones with a clear analytical framework are available (see different workshops held in the recent period under the auspices of the EC, of the European Association of Research Infrastructures in 2012 and 2013 or of the OECD in 2015). Particularly noticeable are the case of a single-site RI carried out for instance on the STFC runned Synchrotron Radiation Source, extending until the end of the RI life (SFTC 2010), and an even more detailed and multidimensional analysis provided on various cases, following the FenRiam approach (see above), among which is the case of the FERMI@ Elettra Free Electron Laser (Péro & Rochow 2010). Beside the tracking of various indicators, the study includes an attempt to run an input-output analysis on the local impact of the RI. Among others, this is also the case of the DESY-HERA facility on which different approaches have been implemented, basically mixing input-output calculation at local level and collection of various data about patent and licensing, collaborations, etc (see for instance Krell 2009). Elaboration of some cost-effectiveness ratio combined with citation statistics at a global European level is also proposed in Del Bo (2016). Besides there are many case studies which include various impact evaluation, especially as regards the impact on the local economy, but not on a similar scope and using methods with highly different levels of scientific quality and technical complexity (most well-known being the CERN case and, ex ante, the European Spallation Source one).

In this context, the EvaRIO project attempted to provide a new comprehensive framework and a set of tools for evaluating the economic impact of RIs, that could complement and add to the above mentioned studies. The focus was on the actors involved in the RI related activities, and relied on the assumption that S&T knowledge creation in general and knowledge creation through RIs in particular resulted from a cumulative and interactive learning process. Only a fine and detailed analysis at the micro-level could properly take those dimensions into account. Part of it was an adaptation of the existing BETA evaluation method that had been extensively used for an ex post evaluation of some of the economic impacts of a large variety of R&D programmes launched by public authorities to support research and innovation (Bach et al., 2003).

Beyond the direct effects of those R&D activities (i.e. their results and the use of their results, corresponding to what was more or less targeted when these activities were designed and launched), the focus of previous BETA evaluation work had always primarily been on the so-called indirect effects, which are the heart of the knowledge creation and diffusion dimensions in the innovation processes. In this BETA approach, indirect effects corresponded to the exploitation of various types of knowledge, networks, modes of organisation, etc. developed or acquired by a given actor during these R&D activities but not directly related to the objectives of these R&D activities. Basically, this approach consisted in identifying (and measuring the impact of) various types of learning processes triggered by the participation in R&D.

The EvaRIO project then aimed at drawing a comprehensive mapping of the different effects that can be attributed to an RI, by taking into account and/or identifying the relevant types of activities related to

Tab. 1 Typology of effects in EvaRIO approach

	RI OPERATOR(S)	RI SUPPLIERS	RI USERS
Direct effects	volume of activities corresponding to the building and operating of RI	volume of activities corresponding to the supplying of resources open as RI	• volume of activities corresponding to the research projects using RI • *direct advantage from using the RI*
***Capacity effects* (capacity : assets + capacity to mobilize and make them evolve)**	change in the capacity due to the operating of the RI, in the field of S&T, Network, Organisation & Methods, Reputation, Human Capital	change in the capacity due to the supplying of resources to the RI, in the field of S&T, Network, Organisation & Methods, Reputation, Human Capital	change in the capacity due to the use of the RI, in the field of S&T, Network, Organisation & Methods, Reputation, Human Capital
***Effects on performance* of RI-related activites**	exploitation of the capacity for enhancing the perfomance as operator of the RI	exploitation of the capacity for enhancing the perfomance as supplier of the RI	exploitation of the capacity for enhancing the perfomance as user of the RI
Indirect effects	exploitation of the capacity for generating economic benefit for the actor "out of RI" : • same research field of actor but not on RI • in other field of research of actor • downstream market/society applications	exploitation of the capacity for generating economic benefit for the actor "out of RI" : • same research field of the actor but not on RI • in other field of activity of the actor • downstream market/society applications	exploitation of the capacity for generating economic benefit for the actor "out of RI" : • same research field of the actor but not on RI • in other field of research of the actor • downstream market/society applications

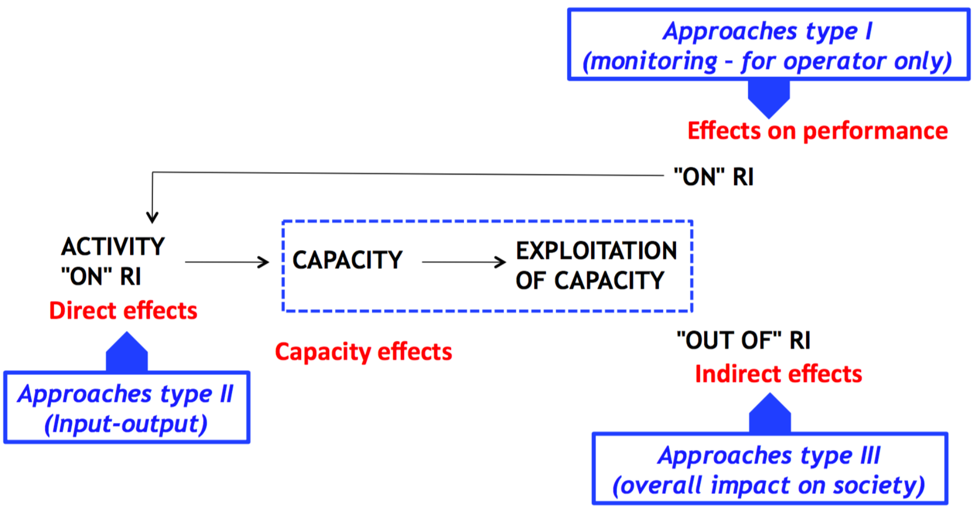
RIs and at the source of possible impacts (setting up, operating, using, etc. the RI), the relevant types of actors concerned (builders, operators, researchers accessing the RI, rest of society: government and funding organisations,...), and the time dimension, i.e. the dynamics of the RI, including possible enrichment of the RI via feed-back loops from users.

The approach distinguishes four broad families of effects (direct, capacity, performance and indirect effects) to be identified and evaluated ex post at the level of interacting individual actors who play different roles (typically operators, suppliers and users) in RI-related activities along the "RI life-cycle" (see Tab. 1).

For each sub-category of effects, the project designed a coherent architecture of specific methods, metrics and indicators with the corresponding indications of information sources and data collection processes. Capacity effects and their exploitation via indirect effects are particularly detailed and are broken down into categories corresponding to different families of knowledge and competences (Science & Technology, Network, Organisation & Method, Reputation, Human Capital).

The evaluation is based on a micro-economic approach, and on interviews / desk data collection methods; economic effects are identified, evaluated at the level of a sample of RI partners, and then if possible aggregated. Correspondingly, only the effects affecting the partners covered are evaluated. In that respect, the evaluation is clearly and purposely limited. However, the framework provides articulation and connections to three families of approaches, in as much as EvaRIO categories of effects can be used as starting points for the latter (see Fig. 1).

Fig. 1 Architecture of effects and links with other approaches



Hence:

- Effects on performance correspond to the classical assessment of the performance of RI operations (in terms of cost, productivity, access, quality, services etc.), related to monitoring activity (type I).
- Direct effects are related to the standard evaluation of the economic impact of building and operating RI, dealing with the so-called direct, indirect and induced effects of RI based on input-output approaches or with various ad-hoc collection of data (type II).
- Indirect effects are connected with the evaluation of societal and economic impacts of scientific achievements obtained thanks to the use of RI as one of the contributors to the scientific advances, e.g. the impact of a new family of drugs or of certain type of therapeutic treatment on mortality, productivity national health expenditure of big pharmas market structures (type III).

Nine case studies were carried out in the Bio-Medical Sciences field: 4 in-depth ones (SOLEIL Synchrotron, EMMA mouse depository network, CERM laboratory and EBI bioinformatics resources), and 5 smaller ones to complement the former on specific points. Not all effects of the typology sketched above were systematically full-scale covered; the case studies allowed to exemplify, test and and refined tools and metrics, and eventually to elaborate guidelines for conducting RI case. Some of the most significant of them will be highlighted.

Of course, the approach also shows some drawbacks (time consuming exercize, difficulty for tackling aggregation and additionality issues…) and it is limited in terms of scope of relevance (evaluation limited to participants, ex-post view, …). Some further developments (for instance as regards metrics) are also required to gain full benefit from some theoretical and methodological choices.

In parallel, the completion of four Focus studies largely based on the case studies led to promising results, especially on the analysis of the open source properties of the dynamics of knowledge creation via RIs, on the mechanisms by which a large RI maintains and manages its flexibility in a science intensive and uncertain environment while contributing to eco-system of RIs for the benefit of the scientific community, and most of all the role of RIs as catalysers of research collaborations and network brokers, analysed thanks to an original use of social network modelling. Recent extensions of this line of work are currently run.

REFERENCES

Arnold E. (2012): "Understanding long-term impacts of R&D funding: The EU framework programme". Research Evaluation (21), pp.332–343.

Bach, L., M.-J. Ledoux, M. Matt (2003): "Evaluation of the BRITE/EU-RAM program", in Shapira P. & Kuhlmann S. (Eds): Learning from science and technology policy evaluation: experiences from the United States and Europe, Nothampton, Massachusetts (USA): Edward Elgar Publishing.

Clarke, S., M. Mawhinney, R. Swerdlow, D. Teichmann (2013): "Project Preparation and CBA of RDI Infrastructure Projects", JASPERS Knowledge Economy and Energy Division, Staff Working Papers, April.

Curaj A. and Pook K. (2011): "FenRiam – Foresight enriched Research Infrastructure Impact Assessment Methodology", produced as part of the 'Research Infrastructures: Foresight and Impact'(RIFI) project co-funded by the European Commission.

Dougan-McCallie C. (2014): "STFC Impact Framework & Evaluation Strategy", retrieved on 5 May at : https://www.stfc.ac.uk/files/impact-publications/stfc-impact-framework-and-evaluation-strategy/

Del Bo C.F. (2016): "The rate of return to investment in R&D: The case of research infrastructures", Technological Forecasting and Social Change – in press, doi:10.1016/j.techfore.2016.02.018

EC (2014): "Guide to Cost-Benefit Analysis of Investment Projects Economic appraisal tool for Cohesion Policy 2014-2020", DG for Regional and Urban policy REGIO DG 02-Communication.

EC (2010): "A vision for strengthening world-class research infrastructures in the ERA", Bruxelles: European Commission, EUR 24186

EIB (2013): "The Economic Appraisal of Investment Projects at the EIB", retrieved on 5 May at: http://www.eib.org/attachments/thematic/economic_appraisal_of_investment_projects_en.pdf

Florio M., S. Forte, C. Pancotti, E. Sirtori and S. Vignetti (2016), « Exploring Cost-Benefit Analysis of Research, Development and Innovation Infrastructures: An Evaluation Framework « , University of Milan - CSIL Working Paper N. 01/2016.

Georghiou, L., Rigby, J. & Cameron, H. (2002). Assessing the Socio–Economic Effects of the EC RTD Framework Programme –ASIF project, CE, retrieved on 5 May at: https://ec.europa.eu/research/evaluations/pdf/archive/other_reports_studies_and_documents/assessing_the_socio_economic_impacts_of_the_framework_programme_2002.pdf

Jordan, G. & Ruegg, R. (2007). Overview of Evaluation Methods for R&D Programs - A Directory of Evaluation Methods Relevant to Technology Development Programs, U.S. Department of Energy Office of Energy Efficiency and Renewable Energy.

Krell, U. (2009). Long-term Sustainability of Accelerator-Driven Research Infrastructures in the Knowledge-Based Economy, retrieved on 5 May at: http://ec.europa.eu/research/infrastructures/index_en.cfm?pg=impact_studies.

MioiR and Technopolis (2012): "Evaluation of Innovation Activities. Guidance on methods and practices", Study funded by the EC-Directorate for Regional Policy, Evaluation Unit C4.

Péro, H. (2009): "Role of RI in a competitive knowledge economy: Lessons learnt from the EC RI actions, retrieved on 5 May at: http://ec.europa.eu/research/infrastructures/pdf/pero_ri-impacts.pdf.

Péro, M. & Rochow, R. (2010). Socio – Economic Impact Report, Free Electron Laser FERMI@Elettra, Brussels: European Commission.

SFTC (2010): "New Light on Science: The Social & Economic Impact of the Daresbury Synchrotron Radiation Source 1981–2008". Science & Technology Facilities Council.

Technopolis (2010). BBMRI: an evaluation strategy for socio-economic impact assessment, The Netherlands: Technopolis Group.

Simmonds P. et al (2013), "Big Science and Innovation", report from technpolis Group for Department of Business, Innovation and Skills, Ref: BIS/13/861

Zuijdam, F. et al. (2011): "The role and added value of large-scale RI", Ministry of Education, Culture and Science (NL), retrieved at: http://www.technopolis- group.com/site/downloads/index.htm.

AUTHORS

BACH LAURENT
BETA (Bureau d'Economie Théorique et Appliquée),
University of Strasbourg/CNRS UMR7522
61, avenue Forêt Noire
67085 Strasbourg
France
Tel +33 3 68 85 21 03

WOLFF SANDRINE
BETA (Bureau d'Economie Théorique et Appliquée),
University of Strasbourg/CNRS UMR7522

UNPACKING THE OPENNESS OF OPEN EVALUATIONS

CHRISTIAN VOIGT, URSULA HOLTGREVE AND ALEXANDER DEGELSEGGER

1 INTRODUCTION

The proposed paper discusses open evaluation in the context of two fields of knowledge-intensive "production" that are in varied ways and to varied degrees based on "openness" themselves: *scientific publishing* and *peer2peer production (maker movement)*.

Scientific publishing traditionally represents the ideal-typical public domain of knowledge sharing that Merton has described as 'communist': "The substantive findings of science are a product of social collaboration and are assigned to the community. ... The scientist's claim to 'his' intellectual 'property' is limited to that of recognition and esteem." (Merton, 1973, p. 273). This accessibility of scientific results and knowledge ideally enables further knowledge generation: The " commons is the 'seedbed' for any production of creative work and innovation: producers of knowledge and cultural goods inevitably stand 'on the shoulders of giants' or at least those of other creative people." (Holtgrewe, 2005). Maker spaces and communities, provide public access (or access based on club membership) to advanced digital manufacturing technology. They bring the immaterial character of open and commons-based production to material goods (Le Roux, 2015). Manufacturing can be individualized, decentralized and opened to amateurs and hobbyists or to subsistence production for one's own needs. Designs and knowledge can be shared and adapted freely. Hence, both fields of production, with different histories, material and immaterial bases and modes of organisation operate on openness, sharing and collaboration in public or semi-public domains. They thus fit in a tradition of knowledge-based, creative and collaborative work and innovation in which openness is not just a mechanism but – to varying degrees – a norm and a programme emphatically aiming at democratization, progress and empowerment.

The term 'open evaluation' is still relatively new and no widely accepted definiton has been adopted. The general approach seems to start with other areas emphasising 'openness', such as the ones already mentioned, deriving a working definition from known principles within these areas. For example, Chesbrough et al.'s (2006) definition of open innovation as "*... the use of purposive inflows and outflows of knowledge to accelerate internal innovation, and to expand the markets for external use of innovation, respectively*" is adapted by Haller (2013), who defines open evaluation as "*... the integration of stakeholders outside the usual group of decision- makers into the assessment of pre-developmental products or services by means of IT- supported acquisition, aggregation and assimilation of quantitative or qualitative judgments*". Clearly, Haller adopts the open innovation maxim that organisations should garner knowledge outside their own boundaries, in this specific case for the purpose of product evaluation. What is interesting in Haller's definition is that it emphasises the inflow of knowledge but neglects a possible outflow, e.g. the availability (openness) of evaluation data and comments outside the specific evaluation effort. In this sense open evaluation resembles crowdsourcing, defined as "*... a collaboration model enabled by people-centric web technologies to solve individual, organizational, and societal problems using a dynamically formed crowd of interested people who respond to an open call for participation.*" (Pedersen et al., 2013). In either case, an open process does not necessarily imply an openly accessible result. As in commercially oriented crowdsourcing, it is possible that no property rights or even information about other contributions are given back to the crowd (Voigt, Havlik, & Hannes, 2013). As we can see from this introductory discussion, unpacking the term 'open evaluation' is likely to generate a conceptually rich picture of considerations to support designers of open evaluations who aim for an open evaluation approach. More concretely, the paper will discuss benefits and risks in open evaluations and point towards an evolutionary process in refining and consolidating open evaluation principles as a result of open collaboration.

The paper starts with the outline of a rough evolutionary perspective that provides a heuristic to compare various "open" processes of knowledge production, providing a meta-. We then outline the notion of open evaluation and explore two 'open evaluation' examples in order to illustrate the current multifacetedness of open evaluation. For a more theoretically informed exploration, we link open evaluation with theory-driven evaluation, participatory evaluation and Weick's (1989) description of theory building as disciplined imagination.

2 AN EVOLUTIONARY PERSPECTIVE ON OPENNESS

In the following we suggest an evolutionary perspective to compare and learn about the different degrees of openness in evaluations. More specifically, we suggest a process to evaluate evaluation processes based on evolutionary principles of variation, selection and retention. Evolutionary processes, according to organisation theorist Karl Weick (1979), require three mechanisms:

1. one generating variation,
2. a selection mechanism, and
3. a mechanism of retention that assures the reproduction of the units that have 'won' the selection.

Proponents of an evolutionary epistemology point to evolution as a knowledge process (Campbell 1988). In this perspective, evolution is fundamental for learning, thinking or science. Evaluation in general formalises and organises a part of these processes, in particular selection, in order to assign resources or attention in legitimate ways. Evaluation processes thus are formalised knowledge processes themselves, in which their own evolutionary mechanisms are applied: alternative evaluation arguments (*variations*) compete for wider acceptance (*selection*) and need to proof their sustainability (*retention*) under different conditions - budgets, complexity of the program evaluated, access to the crowd. The main benefit of an evolutionary perspective on openness could be a deliberate increase in variation as well as more variation in interpreting the pros and cons of different levels of openness. Following a number of additional issues not yet raised but likely to influence degrees of openness.

3 CURRENT DISCUSSIONS OF 'OPEN EVALUATION'

Etymologically, evaluation refers to the action of appraising (or valuing) a certain phenomenon. We are interested in evaluation as a deliberate assessment of a certain phenomenon, which involves planning, process management and a set of stakeholders. Several definitions of evaluation exist depending on the subject matter (science policy, labour market, product development, etc), the type of phenomenon that is evaluated (project, programme, institution or policy evaluation) as well as the applied methodology and design (ex ante, interim, ex post impact evaluation) (Pichler, Sheikh, Jörg, Schuch, & Salhofer, 2013). In the context of the evaluation of research and technological development, the European Commission defined evaluation as *"a systematic and objective process that assesses the relevance, efficiency and effectiveness of policies, programmes and projects in attaining their originally stated objectives. It is both a theory- and practice driven approach"* (European Commission, 2002). In development cooperation, the OECD has defined evaluation as a *"systematic and objective assessment of an on-going or completed project, programme or policy, its design, implementation and results. The aim is to determine the relevance and fulfilment of objectives, development efficiency, effectiveness, impact and sustainability"* (OECD, 2010).

Both definitions stress that evaluation has to be systematic and aim for objectivity, which supports reproducibility and communication of results. Both definitions distinguish evaluations at project, programme and policy level. However, the term evaluation is also sometimes employed to describe the ex ante assessment of proposed activities, which are not necessarily projects (as in the case of design competitions), or the ex post assessment of individual outputs, not necessarily projects either (like a journal paper). So far, it is these types of evaluation in a wider sense where open evaluation has been discussed under this very label.

3.1 OPEN PEER REVIEW OF SCIENTIFIC OUTPUT

Our first example concerns the post publication peer review procedures for scientific outputs. According to Kriegeskorte (Kriegeskorte, 2012), the two major functions of the academic publishing system are to provide access to scientific knowledge and to evaluate scientific outcomes. The evaluation of publications fulfils the function of a quality review process, which, in theory, helps readers to invest their limited time in reading high quality publications. However, evaluation also steers the attention of the scientific community and affects scienceinternal processes as well as the potential impact of scientific results through the citations a publication receives. Conventional peer-review is based on the expertise ascribed to evaluators who are recruited through networks of reputation. Through the anonymity entailed in double-blind reviewing, this personal character of reputation is then in effect neutralised – a fragile social process hampered by mutual subjective investments in scientific authorship, temptations to de-anonymise it, or in fact the impossibility to effectively render more and more specialist project histories anonymous to a well-informed audience. Hence, the (closed) standard of peer review as currently organised (double-blind with neither author nor reviewer knowing each other's identity) has been criticised as non-transparent, based on too few arbitrary judgements and delaying publication (Shanahan & Olsen, 2014). It has also been argued that journal prestige affects evaluation results and that the current system reaches its limit as the burden for the (usually unpaid) evaluators gets too high (ibid.).

Various approaches to open evaluation or open peer review are expected to mitigate these flaws (Ford, 2013). Openness is introduced in this debate with a number of meanings: open can refer to the non-anonymous peer review (single-blind or completely open). It can also refer to community peer review procedures where readers rate the article's excellence in a number of areas (the backing of the argument presented or the significance of the contribution). Among other things, this is expected to render the process more transparent, faster and more democratic. A comparison of the quality of reviewer reports showed that open peer-review yielded generally higher quality than closed peer reviews (Kowalczuk, Dudbridge, Nanda, Harriman, & Moylan, 2013). The idea that reviews become citable comments and can contribute to the reviewer's reputation may be conducive to such results.

Since the aim of this paper is to discuss 'open evaluation' as a more general approach we see the following aspects of open peer review that can be transferred to the open evaluation of projects, programmes and policies:

1. **Evaluators**: How do we define openness in open evaluations? Do we refer to knowing who evaluates, respectively, who is evaluated (non-anonymity)? Does it require a process where a larger community can participate? Could it be a mix of both? Or could non-anonymity be optional? How does the signalling of expertise and reputation of evaluators translate to self-recruited crowds?
2. **Evaluation structure**: What are our means of evaluation (in-depth comments, multi-criteria scales and / or weightings)? How should we combine multiple evaluations (a simple average, evaluations weighted by evaluators' citation index, the number of endorsements an evaluation receives from other members of the community)?
3. **Participation rates**: Non-anonymity requires the registration of potential reviewers and multi-criteria ratings take potentially more time than a simple like in Facebook. People's pervasive time constraints can lead to a lack of in-depth reviews and relatively low numbers of ratings, which affects their reliability. How do we present current evaluations including the degree of evaluators' interest they could attract so far?

Additionally, all three aspects depend on voluntary actions and therefore on cultural and systemic incentives which are most likely specific to scientific disciplines or evaluation domains in general.

3.2 OPEN COMPETITIONS FOR OPEN DESIGNS

Based on open source developments, the maker movement stands for democratized production, with invention and design following a growing

'Do It Yourself (DIY) spirit' (Gershenfeld, 2008). Today rapid prototyping is more accessible than ever before due to affordable computer-aided design software, 3-D printing, laser cutting and a knowledge community that is pushing the limits of what can be produced by individuals. For example, sales of goods on ETSY, an e-commerce marketplace specializing in crafts and maker products, reached a turnover of about 2.4 billion USD in 2015. Sharing platforms such as Thingiverse.com currently provide more than half a million 3D models for users to print or mash (https://www.thingiverse.com/about). Another platform, instructables. com, publishes sets of instructions to produce things the DIY way, ranging from Arduino-based magnetic card readers up to smart homes. However, the likelihood that a 3D model or set of build instructions is reused depends on sophisticated search capabilities as well as evaluations concerning form, function and feasibility of the respective model (Kyriakou & Nickerson, n.d.). The difference between evaluating publications and evaluating 3D models is in the final purpose of the evaluation. Generally, several, mutually supporting or controversial publications are read to feed into a further scientific study, proposal or publication. Selecting a 3D-design for actual production is more exclusive: the maker will aim for the 'one best' solution (or very few solutions) for their purpose – unless the purpose is more research-oriented, to study others' designs and ideas to develop something new. For this the prevalent approach is to use metrics such as downloads of maker files or how often a file was re-used (mashed) within another product.

Companies or sharing platforms that want to go beyond mere peer-2peer ratings organise design competitions to evaluate design submission and nominate a group of finalists and usually a set of winners for different categories. For example instructable.com organises competitions to create open source lab equipment (http://www.instructables.com/contest/buildmylab/) or objects including 3D printed parts (http://www.instructables.com/contest/3dprinting2016/). In both instances, everyone can submit ideas and winning entries are required to adopt some form of Creative Commons license. However, even though the evaluation result is openly accessible, the evaluation procedures are determined by the sponsors of the competition in question. In the case of the 'lab equipment competition', the community can vote on specific entries but these votes are not used for determining the winner, whereas the competition including 3D-printed objects makes explicit use of the community's votes. The latter approach is also typical for idea competitions sponsored by large companies who look for new consumer products, graphic design or other creative input, but unlike the instructables.com competition, here the winning design or idea does not become part of the commons but is owned by the sponsor. An analysis of several open evaluation competitions has highlighted possible tensions between sponsors of a competition and the evaluating community, when both parties prioritize different evaluation criteria (e.g. feasibility of a design versus its originality) or between different community groups (e.g. ultimately the design with the largest network wins rather than an objectively better design) (J. B. Haller, Bullinger, & Möslein, 2011).

Again, the above discussion provides a number of aspects that should inform the use of open evaluations on projects, programmes and policies levels:

1. **Licensing regulations**: Openness in open design competitions can refer to a number of points within an evaluation design, such as (a) the competition being open to all participants, (b) a wider community can evaluate submitted designs and (c) the winning design is openly available (for re-use etc.) to all. Particularly (c) is a matter of licenses and patents attached to a design or the object, and this is where community platforms and commercial crowdsourcing deviate the most.
2. **Sponsor's dominance**: Evaluations and competitions are resource-dependent activities that assign the power of choice to the sponsor. In order to meet the objectives that caused an organisation to sponsor a competition, the sponsor is likely to both feel the need to influence or 'correct' the course of an open evaluation and the 'right' to do so on the intuition that who pays the piper calls the tune.
3. **Biased evaluation due to network effects**: The ICT-based evaluation environment might overly influence an open evaluation or competition due to its technical/social choices. Whereas earlier platforms and review systems base their weighting or rating of evaluators on their own reputation mechanisms, more recent solutions are meshed with other platforms: choices can include one-click registration through existing Facebook accounts or Google Plus accounts (favouring account holders) as well as sharing options through Twitter or LinkdIn (favouring those with large followership).

3.3 DIFFERENT RATIONALES BEHIND OPENNESS

These considerations provide evidence for some limitations of the genera programmatic of openness. In the view of both modernist and Marxian optimists openness empowers creative labour, challenges the institutions of property, markets and capital and enables collaboration, community and creative self-expression. Andrè Gorz writes: "Transforming capital opens up the perspective towards a society of knowledge and culture but resists its development in order to retain its power." (Gorz, 2002, p. 28, translation by authors). For Finnish innovation theorist Ilkka Tuomi open source software represents a genuinely modern convergence of technical and communicative rationality in which Habermas' "forceless force of the better argument" is aligned with technological optimisation: "The culture of hacking is probably the most perfect and frictionless implementation of modernity. [...] As long as it builds itself around those technological artefacts that it produces, it is able to avoid many of those conflicts that make similar efficiency difficult in broader social contexts" (Tuomi, 2002, p. 214). Alternative rationales behind openness may include:

1. the participatory and democratic logic of potentially including amateurs, and everyone
2. the logic of rapid technical or functional optimization (crowds producing "better" solutions)
3. and the logic of resource economy or frugality (crowds producing cheaper solutions).

We suggest that these logics may be articulated in varied combinations and with shifting relevance, and that some trade-offs or dilemmas may be found between them.

3.4 OPENNESS DOES NOT AUTOMATICALLY IMPLY INCLUSIVENESS

Technically-enabled communities tend to be more socially homogeneous than their promotors would like, often consisting of large majorities of young(ish) men with technical backgrounds from wealthier countries

(Lin, 2005; Ghosh, Robles, Glott, 2002). Sociologists of knowledge point out that accessibility to knowledge alone does not guarantee inclusiveness: knowledge is articulated in its utilization (Callon, 1994; Håkanson, 2002) and to be used, requires a heterogeneous ensemble of knowledge goods, expertise embodied in human brains, intersubjective sensemaking processes requiring attention, selection and understanding (Weick, 1995), and communities and networks of practice (Lave, Wenger, 1991). Thus, through its contextuality that is not universally accessible, knowledge is exclusive by definition (Callon, 1994).

3.5 OPENNESS AS A HIDDEN WAY OF EXPLOITATION

More recent theorists of modernity and network capitalism add a further skeptical note to the discussion of networks and commons-based peer production. Boltanski, Chiapello (2002) argue that contemporary capitalism has absorbed the norms and values of networked collaboration into intensified and globalized production and exploitation, and theorists of network and platform capitalism point out that apart from formal wage labour, hybrid modes of work, play, creative expression and volunteering can also feed into value creation that is appropriated by the owners of networks and platforms (Scholz, 2012; Scholz, Liu, 2011; Lanier, 2013). The democratic, participatory and empowering impetus of "openness" may thus be absorbed into a further round of exploitation and self-administered discipline.

4 LINKS WITH EXISTING APPROACHES TO EVALUATIONS

The purpose of this literature review is to discuss the meaning of 'open evaluation' in the context of established approaches to evaluation. The previous discussion of open evaluations in the context of open peer review and open design contests has already highlighted a number of aspects to take into account when planning for an open evaluation. In the following three sections we revisit evaluation as (a) an output- or theory-driven process; (b) a stakeholder-driven or participatory process or (c) an evolutionary learning process.

By investigating these areas we hope to gain awareness of evaluation-related discussions[1], taking place under a different label but still being relevant and informative to address the issues identified in the previous section.

4.1 THEORY-DRIVEN EVALUATIONS

An issue with open evaluations identified earlier refers to a lack of authoritative power, when evaluations are based on the judgements of a (possibly non-specialised) community. Lay-people would lack the technical sophistication compared to domain experts, so the argument. Hence, assuming that domain expertise is to a large part captured in theories, we look at theory-driven evaluations as a field with a potential to inform open evaluations. Theory-driven evaluations are, just as objectives-based evaluations, quasi-evaluations as they are limited to the objectives anticipated by a program and tend to refrain from analysing actual needs -and therefore the program's legitimation - or side-effects of the program (Stufflebeam & Coryn, 2014). Theory-driven evaluations require sufficiently detailed and tested theories describing the evaluation area (e.g. higher education institutions). If such a theory exists, evaluators can choose the relevant variables and causal relationships in order to determine what worked and what didn't work and why it didn't work, eventually recommending actions for improvement. However, a review of theory-driven evaluations showed that evaluations often apply reductionist models that only insufficiently capture the complexities of the field to be evaluated (Coryn, Noakes, Westine, & Schröter, 2011). However, if validated theories exist, theory-driven evaluations can deliver timely and to-thepoint recommendations to the program sponsors. In the context of open evaluations, this would not necessarily be strict alternatives, as domain experts and so called 'lay-people' can co-produce evaluations. Such an approach might, however, benefit from a reconceptualization of lay-people's knowledge, differentiating between expertise and experience (Collins & Evans, 2002). Extended experience with a program counts in so far as it can provide insights into details or long-term patterns not known to more theoretically informed experts. Similar reasoning has also led to participatory evaluations described in the next section.

4.2 PARTICIPATORY EVALUATIONS

In the 1990s, discussions around the concept of participatory evaluation moved beyond the principles of stakeholder-centred evaluation to involve more aspects of what would now be called open evaluation. Participatory evaluation approaches recognise the relevance of stakeholder and user knowledge. They call for an opening of the evaluation process to include non-expert-based forms of knowledge. (Cousins & Whitmore, 1998) distinguished practical participatory evaluation and transformative participatory evaluation. While the former "supports organizational and program decision making and problem solving" (ibid., 6), the latter includes an emphatic empowerment approach and a commitment to democratization of the phenomenon in question. Cousins and Whitmore (ibid.) characterise participatory forms of evaluation along three dimensions: first, the control of the evaluation process (from researcher/evaluator-controlled to practitioner-controlled); second, the stakeholder selection for participation (from primary users such as program managers to 'all legitimate groups', especially beneficiaries); third, the depth of participation (from consultation to deep participation). Practical participatory evaluation, then, is characterised by balanced process control, primary user participation and deep participation. Transformative participatory evaluation combines practitioner control with, participation by all legitimate groups and deep participation. With regard to our discussion of openness in open evaluations, the literature on participatory evaluation sheds light on two dimensions where openness can be brought into the evaluation process: first, which stakeholders should be involved, and, second, the format of engagement. As to the purpose that the openness serves, we find similar implicit or explicit expectations in the literature for citizen science (Dickinson et al., 2012) and participatory technology assessment (Joss und Bellucci 2002; Degelsegger und Torgersen 2011): stakeholder involvement and participatory evaluation approaches are discussed as functional (supporting decision-making, enhancing the use of evaluation findings, improving the effectiveness of the intervention

1 It's important to note that the ensuing discussions of theory-driven and participatory evaluations are all on the level of program and policy evaluations

that is evaluated) or as value-driven (democratising the production and use of relevant knowledge (Brisolara, 1998).

5 CONCLUSION

The paper discussed openness in open evaluation adopting an evolutionary perspective on knowledge in general and variation and selection of knowledge more specifically. The papers main contribution includes a discussion of 'open evaluations' used for open peer reviews and open design competitions as well as related issues such as rationalisations, inclusiveness and the possible misappropriation of open evaluations.

Having explored several facets of open evaluation more in-depths we reviewed established evaluation approaches in order to revisit issues of evaluation expertise, engagement of evaluators and socio-economic implications of evaluations (e.g. evaluation data as commons or open licenses for submissions to open competitions). 1 It's important to note that the ensuing discussions of theory-driven and participatory evaluations are all on the level of program and policy evaluations (e.g. evaluation data as commons or open licenses for submissions to open competitions).

Finally, we do not conclude that there must always be a maximum of openness in open evaluations but would like to see a more differentiated and widely shared learning process as to varying implementations of open evaluations. A more detailed understanding of openness can only be a first step, empirical work is now needed to understand how much openness is requested, implemented and, eventually, whether it improved the usefullness of the evaluation for clients and affected communities alike.

6 REFERENCES

Benkler, Y. (2006) The wealth of networks: how social production transforms markets and freedom (New Haven Conn: Yale UP).
Boltanski, L. and Ève Chiapello (2002) The New Spirit of Capitalism. Paper presented to the Conference of Europeanists, March 14-16., 2002, Chicago,http://www.sociologiadip.unimib.it/mastersqs/rivi/boltan.pdf.

Brisolara, S. (1998). The History of Participatory Evaluation and Current Debates in the Field. New Directions for Evaluation, 80, 25–41.

Brown Urban, Jennifer / Hargraves, Monica / Trochim, William M. (2014): Evolutionary Evaluation: Implications for evaluators, researchers, practitioners, funders and the evidence-based program mandate, in: Evaluation and Program Planning, 45, 127-139.

Callon, M. (1994) 'Is Science a Public Good?', Science, Technology & Human Values, 19, 4, 395–424.

Campbell, Donald T. (1988): Evolutionary epistemology, in: Overman, E.S. (Ed.): Methodology and epistemology for social science: Selected papers of Donald T. Campbell, Chicago: University of Chicago Press.

Chesbrough, H., Vanhaverbeke, W., & West, J. (2006). Open innovation: Researching a new paradigm. OUP Oxford.

Collins, H. M., & Evans, R. (2002). The third wave of science studies studies of expertise and experience. Social Studies of Science, 32(2), 235–296.

Coryn, C. L., Noakes, L. A., Westine, C. D., & Schröter, D. C. (2011). A systematic review of theory-driven evaluation practice from 1990 to 2009. American Journal of Evaluation, 32(2), 199–226.

Cousins, J. B., & Whitmore, E. (1998). Framing Participatory Evaluation. New Directions for Evaluation, 80, 5–23.

Degelsegger, A., & Torgersen, H. (2011). Participatory paternalism: citizens' conferences in Austrian technology governance. Science and Public Policy, 38(5), 391–402.

Dickinson, J. L., Shirk, J., Bonter, D., Bonney, R., Crain, R. L., Martin, J., … Purcell, K. (2012). The current state of citizen science as a tool for ecological research and public engagement. Frontiers in Ecology and the Environment, 10(6), 291–297.

European Commission. (2002). RTD Evaluation Toolbox - Assessing the Socio-Economic Impaft of RTDPolicies (IPTS Technical Report Series).

Ford, E. (2013). Defining and Characterizing Open Peer Review: A Review of the Literature. : Journal of Scholarly Publishing, 44(4), 311–326. http://doi.org/dx.doi.org/10.3138/jsp.44-4-001

Gershenfeld, N. (2008). Fab: the coming revolution on your desktop--from personal computers to personal fabrication. Basic Books.

Ghosh, R. A., Gregorio Robles and Rüdiger Glott (2002) Free/Libre and Open Source Software: Survey and Study Part V,http ://floss1.infonomics.nl/finalreport/FLOSSfinal-5.pdf, Download 25.7.2002.

Gorz, A. (2002) 'Welches Wissen? Welche Gesellschaft?' in Heinrich-Böll-Stiftung (ed.) Gut zu wissen. Links zur Wissensgesellschaft (Münster: Westfälisches Dampfboot).

Håkanson, L. (2002) Creating Knowledge - The Power and Logic of Articulation. (What the Fuss is All About).

Paper prepared for the 2002 LINK Conference, November, 1-2, 2002,http://www.cbs.dk/departments/int/link2002/papers/Hakanson.doc.

Haller, J. (2013). Open evaluation: Integrating users into the selection of new product ideas. Springer Science & Business Media.

Haller, J. B., Bullinger, A. C., & Möslein, K. M. (2011). Innovation contests. Business & Information Systems Engineering, 3(2), 103–106.

Holtgrewe, U. (2004) 'Articulating the Speed(s) of the Internet: The Case of Open Source/Free Software', Time & Society, 13, 1, 129–146.

Holtgrewe, U. (2005) 'Intellectual Property, Communism and Contextuality. A Non-Essentialist Exploration of German Digital Copyright and the Public Domain', Science, Technology and Innovation Studies, 1, 1, 39–58.

King, J. A. (2007). Making Sense of Participatory Evaluation. New Directions for Evaluation, 114, 83–86.

Kowalczuk, M. K., Dudbridge, F., Nanda, S., Harriman, S. L., & Moylan, E. C. (2013). A comparison of the quality of reviewer reports from author-suggested reviewers and editor-suggested reviewers in journals operating on open or closed peer review models. F1000 Posters, 4, 1252.

Kriegeskorte, N. (2012). Open evaluation: a vision for entirely transparent post-publication peer review and rating for science. Frontiers in Computational Neuroscience, 6(79), 1–16. http://doi.org/doi: 10.3389/fncom.2012.00079

Kyriakou, H., & Nickerson, J. V. (n.d.). Novelty and Reuse in an Open Innovation Community.

Lanier, J. (2013) Who owns the future? (New York: Simon & Schuster).

Lave, J. and E. Wenger (1991) Situated learning - Legitimate Peripheral Participation (Cambridge: Cambridge UP).

Le Roux, S. (2015) 'The intangible economy: FabLabs "inividualised production of objects" A stage in liberating the function of innovation', Journal of Innovation Economics and Management, 2, 99–116.

Lin, Y. (2005) Inclusion, diversity and gender equality: Gender dimensions of the Free/Open Source Software Development,http://opensource.mit.edu.

Merton, R. K. (1973) 'The Normative Structure of Science' in N. Storer (ed.) The Sociology of Science (Chicago: University of Chicago Press).

Morozov, E. (2013). To Save Everything, Click Here: Technology, Solutionism, and the Urge to Fix Problems that Don't Exist. Penguin UK.

O'Mahony, S. (2007) 'The governance of open source initiatives: what does it mean to be community managed?', Journal of Management and Governance, 11, 2, 139–150.

OECD. (2010). DAC's Glossary of Key Terms in Evaluation and Results Based Management. Paris: OECD.

Pedersen, J., Kocsis, D., Tripathi, A., Tarrell, A., Weerakoon, A., Tahmasbi, N., … de Vreede, G.-J. (2013). Conceptual foundations of crowdsourcing: A review of IS research (pp. 579–588). Presented at the System Sciences (HICSS), 2013 46th Hawaii International Conference on, IEEE.

Pichler, R., Sheikh, S., Jörg, L., Schuch, K., & Salhofer, D. (Eds.). (2013). Evaluation Standards in Research and Technology Policy. Vienna. Retrieved from http://www.fteval.at/upload/fteval_Standards_english.pdf

Scholz, T. (2012) Digital labour: the Internet as playground and factory (New York: Routledge).

Scholz, T. and Laura Y. Liu (2011) From mobile playgrounds to sweatshop city,http://situatedtechnologies.net/files/ST7-MobilePlaygrounds_SweatshopCity.pdf.

Sebald, G. (2008) Offene Wissensökonomie. Analysen zur Wissenssoziologie der Free/Open Source- Softwareentwicklung (Wiesbaden: VS).

Shanahan, D. R., & Olsen, B. R. (2014). Opening peer-review: the democracy of science. Journal of Negative Results in Biomedicine, 13(1), 1.

Stehr, N. (1999) '"Wissensgesellschaften oder die Zerbrechlichkeit moderner Gesellschaften"' in W. Konrad, W. Schumm (eds.) Wissen und Arbeit (Münster: Westfälisches Dampfboot).

Stehr, N. (2001) Wissen und Wirtschaften. Die gesellschaftlichen Grundlagen der modernen Ökonomie (Frankfurt/Main: suhrkamp).

Stufflebeam, D. L., & Coryn, C. L. (2014). Evaluation Theory, Models, and Applications (Vol. 50). John Wiley & Sons.

Tuomi, I. (2002) Networks of Innovation. Change and Meaning in the Age of the Internet (Oxford: Oxford UP).

Voigt, C., Havlik, D., & Hannes, L. (2013). Situated, crowdsourced microlearning: From micro-content to micro-arguments. In MicroLearning Conference. Krems, Austria.

Weick, K. E. (1979) Der Prozeß des Organisierens (Frankfurt/Main: suhrkamp).

Weick, K. E. (1989). Theory construction as disciplined imagination. Academy of Management Review, 14(4), 516–531.

Weick, K. E. (1995) Sensemaking in organizations (Thousand Oaks, Calif. u. a: Sage).

AUTHORS

CHRISTIAN VOIGT
Zentrum für Soziale Innovation
Vienna, Austria
voigt@zsi.at

URSULA HOLTGREVE
Zentrum für Soziale Innovation
Vienna, Austria
holtgreve@zsi.at

ALEXANDER DEGELSEGGER
Zentrum für Soziale Innovation
Vienna, Austria
degelsegger@zsi.at

POSSIBLE WAYS OF MEASURING THE UPDATE AND IMPACT OF OPEN SCIENCE

DIETMAR LAMPERT AND MARTINA LINDORFER

INTRODUCTION

Open science (OS) does not only open up new ways of creating and sharing knowledge, of disseminating results of individual components along the research process but of ways of evaluating science more nuanced, fair, and precise. It is no secret that the evaluation does not start at the end; rather, it depends very much of how processes are defined and what measures are applied along the way. In short, it matters how the OS (Open Science) and the diversification or diffusion of research that it brings along is adopted – adopted not just by the scientific research community but by society. This is not only a matter of technological developments but also of changes in cultural practice. It is yet not clear how the uptake and impact of OS practice ought to be monitored and measured, especially societal impacts. This article is based on the results of a study on Open Science conducted for the European Commission.

Expectations of OS impacts are high. As summarised in a recent report from the OECD (2015), the positive factors associated with OS are e.g. improving efficiency in science, Increasing transparency and quality in the research validation process, or increasing the knowledge spillovers to the economy. In the literature review and in the interviews with OS experts that we conducted, there is a general consent that possible new indicators for the monitoring and assessment of scientific production and its impact need to be agreed on by all stakeholder groups, in light of a major redesign of the scientific process provoked by OS. However, there is yet very little substance to build upon, as will be shown later on.

WHAT IS ALREADY BEING MEASURED

There have been several attempts to set new indicators for the successful uptake and impact of OS practices. One such example is the so called open access citation advantage (OECD 2015), meaning the tendency that open access (OA) publications receive more citations relative to non-open access ones. Some studies analyse the correlation between citations counts, publication format (OA or non-OA), and the quality of articles to see if there is a quality advantage or quality bias according to publication format.

Latest at the outset of altmetrics it became clear, that new evaluation systems are needed – evaluation of research is currently based on teaching and bibliometric indicators that do not take into account a whole array of contributions to and resulting from the research process (data, methods, codes, insights, ideas, trainings, participations in all kinds of activities, etc.). One can see that, with altmetrics, the efforts go beyond the traditional citation metric; however, the scope is still fairly limited. The concepts involved in OS exceed that scope by far.

WHAT TO ACTUALLY MEASURE

One of the main objectives of the study underlying this article was to propose a framework for an OS observatory which monitors the progress of OS in Europe on a continuous basis. The indicators suggested in the article shall therefore be useful to monitor the uptake and impact of OS. Also, indicators shall measure if OS practices make science more accessible for a wider audience, whereby Fecher and Friesike (2014:19) see accessibility in the double sense: (a) accessibility of the research process and (b) comprehensibility of the research result. This understanding suggests that the relationship between science and society must be reflected in the indicators in any case.

METHODOLOGY

The initial set of indicators for measuring the uptake and impact of OS is based on a series of methods fully described in the article. Starting with a literature review, the study went on developing future scenarios that brought to light new necessities and application of potential indicators. Following a series of expert workshops, the indicators were tested by a select group of external experts via an online survey.

AN INITIAL SET OF NEW INDICATORS

Resulting from the above-mentioned applied methodology, our first set of new possible indicators for measuring the uptake and impact of OS is structured such that they are categorised into two major groups/dimensions and seven sub-dimensions:

a. the scientific process:
   - conceptualisation and data gathering/creation
   - analysis
   - diffusion of results
   - review and evaluation
b. the system level:
   - reputation system, recognition of contributions, trust

- open science skills and awareness
- science with society

Each of the above-mentioned dimensions entails a cluster of indicators. Those will be presented in the final article in terms of their nature, their relevance, and the stakeholder group responsible for adopting and further developing an indicator. For this extended abstract, we only provide a brief discussion of each dimension.

CLUSTER I: CONCEPTUALISATION & DATA GATHERING/CREATION

Important questions in this dimension are whether the quality of data and information is adequate, e.g. whether the data were properly cleaned, whether they are curated, are metadata provided, etc. Recent policy trends involve mandatory rules and requirements (most commonly, funding agencies mandate public access to funded research), and the development of infrastructure to enable OS. Fewer initiatives relate to non-monetary incentive mechanisms like the definition of new reward/promotion systems. (ibid.)

Scientific work must no longer be restricted to measuring final products (such as articles), but should measure the development of the individual steps of the scientific workflow. Furthermore, results will differ according to disciplines, fields, or data types. Indicators in this dimension cover e.g. research funding organisations requiring the open provision of data/code, the accessibility of data/code, or the availability of metadata.

CLUSTER II: ANALYSIS

Respondents in this cluster argue that open methods contribute to improving the reliability of research results but that the impact of the open methods were still marginal because their use is not spread widely yet in the research community. Indicators in this cluster that are easier to design and monitor are data citations[1] and code/software citations, a possible new one might be content citations.

CLUSTER III: DIFFUSION

We deliberately chose the term "diffusion" (of results) instead of the term most commonly used term in academia: "publication". We want to stress that diffusion can and – some would argue – should start well before the results are in. In our online assessment, several comments underpinned the need to get away from the traditional paper publishing models and find indicators that gauge the growth of dissemination channels other than journals. Participants stated that journals are becoming irrelevant in many fields already. Impact of OS can more easily be captured in those cases where open communication and responsive attitude to feedback has actually changed the trajectory of research, e.g. a sideline turned into the main thing, a bug/design issue was detected, or the project just responded (or even emerged in response) to what is happening in society.

CLUSTER IV: REVIEW AND EVALUATION

Currently, peer review is the standard practice to assure quality of scientific output. Traditional peer review has well known shortcomings, though, such as little credit given to reviewers, lack of transparency and limited verification of scientific results[2]. Open peer review is often mentioned as an alternative, but not without the same amount of criticism. In the Open Science community, however, there is certain agreement that transparency measures need to be taken in the review and evaluation process. A multitude of suggestions is up in the air, some being considered as "incremental", meaning that they would not do much harm to the current review procedure, and others as "radical" or quite transformative. Adding transparency to the review process can happen at various stages of the scientific process and therefore be more or less transformative. One option would be to make grant proposals publicly accessible at various points of time (after the project has ended, along with the final project reports, at the beginning of a project, at the point of announcing funding decisions, upon submission to the funder and during the drafting phase)[3]. Another would be to make the peer review public. This can again happen in an incremental form, meaning that some knowledge within the peer review process is made openly accessible, or in a radical form, meaning that transparency of knowledge becomes a separate pillar of legitimacy itself.[4] Open peer review is currently a highly contested field and so is the choice of respective indicators. This can also be said for the question how societal relevance of research should be treated and assessed in evaluation. A rather easy measure could be to make the "impact statement" of a proposal publicly accessible. A labelling system for expected impact (oriented on e.g. the Sustainable Development Goals) could be an option to create clearer evaluation references.

CLUSTER V: REPUTATION SYSTEM, RECOGNITION OF CONTRIBUTIONS, TRUST

The uptake of OS practice in the research process is unlikely to flourish if researchers fear it is not properly acknowledged and officially recognised. This is underpinned in the initially mentioned surveys on researchers' attitudes towards OS, which reveal low factual progress in putting OS into practice. Reward mechanisms for data sharing are currently especially weak and researchers might choose rather not to spend a serious amount of time in cleaning and curated their data for the re-use of others. Some organisations (datacite, ORCID, Figshare, Dryad Digital Repository, ResearcherID) have propositions for data citation tools which would credit authors for data and metadata sharing, but "in most countries the existing framework does not promote sharing efforts, especially with respect to results, data sets or other research material at the pre-publishing phase"[5]. Formal recognition of a variety of contributions along the scientific process (e.g. to the selection of research topics, formulation of hypotheses, project participations, review activities, etc.) have yet to be adopted.

1 platforms that may provide data on data citation: DataCite, ORCID , Figshare, The Dryard Digital Repository, ReseacherID.
2 OECD:2015:50
3 Mietchen, D, The Transformative Nature of Transparency in Research Funding
4 D., Milanesi, E., Koenig, T.(2014): Grant Application Review: The Case of Transparency
5 OECD:2015:89

CLUSTER VI: OS SKILLS & AWARENESS

OS-related skill development across disciplines will be a crucial factor for the maturation of OS in Europe. Researcher's skills in OS (e.g. curating and maintaining large data sets) differ across disciplines due to different traditions or training opportunities in digital tools and data handling. There is a substantial need for further training of researchers and scientists in handling big, multi-layered and complex data sets. Accordingly, indicators in this cluster cover e.g. the monitoring of skilled personnel, research personnel active in OS, or the awareness and use of open standards.

CLUSTER VII: SCIENCE WITH SOCIETY

This cluster is about finding indicators that assess effects of OS on the promotion of the engagement of citizens in science and research. As Mietchen, Mounce, and Penev (2015) observed, most of the research process is hidden from public view through multiple layers of obfuscation as a heritage of conventions and habits from the paper era. This has begun to change, also because digital technologies enable engagement and popularisation. Popularisation activities are understood as targeting a wide audience and a non-specialised public. Consequently, relevant new indicators gauge, among others, citizens' engagement in (open) science, research communication (beyond academia), or the accessibility of data that are of public interest.

CONCLUSION AND OUTLOOK

Designing indicators to measure the uptake and impact of Open Science is a challenge, not least of all because the concept itself is still evolving. Open Science is necessarily broad because it is composed of many dimensions (e.g. along the scientific research process) and embedded in a larger.

M indicators proposed in this report are new and not gathered/surveyed/evaluated automatically (yet). Consequently, a first vital step is put the necessary mechanisms in place. To achieve this, the stakeholder groups that are primarily involved in/responsible for an indicator[6] are provided in the full article.

An essential precondition for indicators to work as intended is that all concerned stakeholder groups are involved in their design and evolvement. They all need to agree on what an indicator should measure (and what it should) and how it should be used (and what it must not be used for). Furthermore, indicators need to be flexible enough to accommodate differences, e.g. in research fields, and allow the emergence of new developments. The differences in research fields can be considerable, as is the pace at which OS is being adopted. Those differences will need to be elaborated and reflected in the relevant indicators.

Finally, new indicators need to be tested – not just discussed – before being adopted on a larger scale. This can be done in small experiments with individual, selected indicators.

BIBLIOGRAPHIC REFERENCES

Alperin J.P., Babini, D., Fischman, G.E. (2013): Open Access indicators. Assessing growth and use of OA resources from developing regions - the cases of Latin America and Africa, developed by UNESCO, CLACSO, PKP and FLACSO Brasil

Bartling, S., Friesike, S. (2014): Opening Science. The evolving guide on how the internet is changing research, collaboration and scholary publishing, Springer Open

Bornmann, L. (2014): Measuring the broader impact of research: The potential of altmetrics, published by researchgate

Buschmann, Katrin, Stefan Kasberger, Peter Kraker, Katja Mayer, Falk Reckling, Katharina Rieck, Michela Vignoli (2015): Open Science in Österreich: Ansätze und Status (Open Science in Austria: Approaches and status), in Information. Wissenschaft & Praxis 2015; 66(2–3): 1–9. DOI 10.1515/iwp-2015-0025

Delfanti, A., Pitrelli, N. (2015): Open science: revolution or continuity?, In: Open Science, Open Issues, Series UC Davis Previously Published Works

EC (2015): Indicators for promoting and monitoring Responsible Research and Innovation. Report from the Expert Group on Policy Indicators for Responsible Research and Innovation, EUR 26866 EN

Fecher, Benedikt, and Sascha Friesike (2014) „Open Science: One Term, Five Schools of Thought." Opening Science: 17.

Finnish Ministry of Education and Culture (2014): Open science and research leads to surprising discoveries and creative insights. Open science and research roadmap 2014-2017. Reports of the Ministry of Education and Culture, Finland 2014:21

Galloway, Linda M. and Pease, Janet L. (2013): „Altmetrics for the Information Professional: A Primer". Libraries' and Librarians' Publications. Paper 105.

Jensen, P., Jean-Babstiste, R., Kreimer, P., Croissant, Y. (2008): Scientists who engage with society perform better academically, http://arxiv.org/abs/0810.4672

Mietchen D, Mounce R, Penev L (2015): Publishing the research process. Research Ideas and Outcomes 1: e7547. doi: 10.3897/rio.1.e7547

Mietchen, D. (2014): The Transformative Nature of Transparency in Research Funding, Published: December 30, DOI: 10.1371/journal.pbio.1002027

Mingers, John and Loet Leydesdorff (2015): A Review of Theory and Practice in Scientometrics, European Journal of Operational Research

6 designing, measuring, interpreting, and/or adapting an indicator

OECD (2015): "Making Open Science a Reality", OECD Science, Technology and Industry Policy Papers, No. 25, OECD Publishing, Paris. http://dx.doi.org/10.1787/5jrs2f963zs1-en

OECD (2015b): OECD Digital Economy Outlook 2015, OECD Publishing, Paris. DOI: dx.doi.org/10.1787/9789264232440-en

Scheliga, K., Friesike, S. (2014): Putting Open Science into practice: A social dilemma? http://firstmonday.org/ojs/index.php/fm/article/view/5381/4110

AUTHORS

DIETMAR LAMPERT
ZSI – Centre for Social Innovation, Austria
Linke Wienzeile 246, 1150 Vienna
Austria
lampert@zsi.at
+43 1 4950 442

MARTINA LINDORFER
ZSI – Centre for Social Innovation, Austria
Linke Wienzeile 246, 1150 Vienna
Austria

POSSIBLE CAN POLICY DOCUMENTS BE USED AS SOURCES FOR MEASURING SOCIETAL IMPACT? AN EMPIRICAL STUDY BASED ON CLIMATE CHANGE RESEARCH

LUTZ BORNMANN, ROBIN HAUNSCHILD AND WERNER MARX

In recent years, societal impact measurements of academic research have become more and more important. This trend is not only visible by their consideration in national evaluation systems (e.g. the UK Research Excellence Framework), but also in the commercial success of providers delivering altmetrics data (e.g. Altmetric) which propose that altmetric scores can be used to measure societal impact. Currently, the most important and most frequently used method of societal impact measurement is the case study approach in which cases of research are described leading successfully to a specific form of societal impact (King's College London and Digital Science, 2015). However, case studies have the disadvantages that they are expensive, the results are biased towards success stories, and the results for different entities (e.g. universities) are not comparable. Whereas bibliometric indicators have emerged as the most important metrics to measure the recursive impact of research, the development of metrics for the measurement of societal impact is challenging.

This study focusses on a relatively new form of impact data (provided by Altmetric), which could complement Google patent citations (Kousha & Thelwall, in press) and clinical guideline citations (Thelwall & Maflahi, 2015): mentions of publications in policy documents. It is an interesting form of impact measurement compared to other altmetrics (e.g. mentions in tweets and blogs) because (1) it is target-oriented (i.e. it measures the impact on a specific sector of society) and (2) it focusses on a relevant part of society for research – the policy area. Many research topics are policy-relevant (e.g. health care or labor market research) and it is interesting to know in the context of wider impact evaluations which (kind of) publications have more or less impact.

In this study, we us a comprehensive dataset of papers on climate change to investigate a new source of altmetric data: mentions in policy documents. Climate change is particularly useful in this respect because the topic is very policy relevant since many years. Thus, we expect to find a large number of papers mentioned in policy documents in comparison with other research fields – especially because corresponding policy sites are continuously evaluated by Altmetric. However, the results of our analyses are contrary to our expectation: Out of n=191,276 publications on climate change in the dataset, only 1.2% (n=2,341) have at least one policy mention. The rate of 1.2% is also small in comparison with the result of Kousha and Thelwall (in press) who showed that "within Biomedical Engineering, Biotechnology, and Pharmacology & Pharmaceutics, 7% to 10% of Scopus articles had at least one patent citation". The result of this study contradicts the claim of Khazragui and Hudson (2015) that "it is rare that a single piece of research has a decisive influence on policy. Rather policy tends to be based upon a large body of work constituting 'the commons'" (p. 55). The low percentage of 1.2% which we find in this study might be due to the fact that Altmetric quite recently started to analyze policy documents and the coverage of the literature is still low (but will be extended). However, the low percentage might also reflect that only a small part of the literature is really policy relevant and most of the papers are only relevant for researchers studying climate change. Another reason for the low percentage might be that policy documents may not mention every important paper on which a policy document is based on.

In order to find out which kind of papers are more or less interesting in the policy context (e.g. articles or reviews), we compare the distribution of papers among climate change papers (CCP) and climate change papers mentioned at least once in policy documents (CCP_P). The results show that the policy literature tends to cite research which has been published a longer time ago than researchers cite in their papers. Thus, research papers seem to need more time to produce impact on politics than on research itself. As expected, reviews are overrepresented among CCP_P: the observed CCP_P value is higher than the expected value delivered by the CCP distribution. Reviews summarize the results of many primary research papers and connect research lines from different research groups. Good reviews save the labor of reviewing the literature on one's own responsibility. In this study, we further reveal that papers published in Nature and Science as well from the areas "Earth and related environmental sciences" and "Social and economic geography" are especially relevant in the policy context.

This study is a first attempt to study a new source of altmetric data: mentions of scientific publications in policy documents. We encourage that further empirical studies follow because the data source is of special interest in the use of altmetric data for measuring the broader impact of research. It will be interesting to see whether more papers are used in policy documents in upcoming years (because of the wider coverage of the policy literature by Altmetric).

REFERENCES

Khazragui, H., & Hudson, J. (2015). Measuring the benefits of university research: impact and the REF in the UK. Research Evaluation, 24(1), 51-62. doi: 10.1093/reseval/rvu028.

King's College London and Digital Science. (2015). The nature, scale and beneficiaries of research impact: An initial analysis of Research Excellence Framework (REF) 2014 impact case studies. London, UK: King's College London.

Kousha, K., & Thelwall, M. (in press). Patent citation analysis with Google. Journal of the Association for Information Science and Technology. doi: 10.1002/asi.23608.

Thelwall, M., & Maflahi, N. (2015). Guideline references and academic citations as evidence of the clinical value of health research. Journal of the Association for Information Science and Technology, n/a-n/a. doi: 10.1002/asi.23432.

AUTHORS

LUTZ BORNMANN
Division for Science and Innovation Studies
Administrative Headquarters of the Max Planck Society
Hofgartenstr. 8,
80539 Munich, Germany.
E-mail: bornmann@gv.mpg.de
Phone number: +49-89-2108-1265

ROBIN HAUNSCHILD
Max Planck Institute for Solid State Research
Heisenbergstr. 1,
70569 Stuttgart, Germany.
E-mail: r.haunschild@fkf.mpg.de

WERNER MARX
Max Planck Institute for Solid State Research
Heisenbergstr. 1,
70569 Stuttgart, Germany.
E-mail: w.marx@fkf.mpg.de

EXTENSION OF EVALUATION THEORY TREE BASED ON MAPPING KNOWLEDGE DOMAIN

LI XINGRUI AND LI QIANG

INTRODUCTION

Evaluation theory, a.k.a. evaluation method or evaluation model, aims to produce an applicable and feasible strategy or diagnose, by combining, enhancing and practicing multiple evaluation techniques. Such studies draw from investigation of specific problems and cover the whole evaluation process from planning, practicing to result applying. Now that evaluators have increasingly diversified theoretical tools to meet their needs, the new focus of evaluation research is how to form an organized and systematic social studies theoretical framework through proper classification.

House (1978) is one of the pioneers in classification of evaluation theories. He classified evaluation theories into eight kinds of models and exemplified them respectively, according to their features in these five dimensions, including the audiences, what the model assumes consensus on, the methodology of data collection, ultimate outcome expected, and the typical question. Williams (1989) drew from the similarities of 14 researchers' theoretical models and, through cluster analysis, classified them into three groups (Applications Approach, Flexible Approach and Formal Approach) on his theory map, based on their features in the following four dimensions: qualitative vs. quantitative, accountability vs. policy-orientation, client participation vs. non-participation, and general utilization vs. decision-making utilization. Alkin (2004) observed that, the development of evaluation theory could be portrayed as a tree, which roots in a dual foundation of accountability & control and social inquiry. Upon the tree, there are three primary branches, *methods, valuing and use.* The methods branch, a continuation of social inquiry, deals with the "knowledge construction" of evaluation theories; the use branch focuses on how evaluation will support decision-making, in what way evaluation information will be used and who will use it; the valuing branch attach more importance to systematically facilitating the placing of value by others. In this way, Alkin classified leading researchers and placed them on sub branches branching out from primary ones, which consequently form the evaluation theory tree. While Donaldson (2012) believed that, evaluation theories tend to be application-oriented and should evolve with their subjects and focuses, which is why the development of evaluation theories can be depicted as rivers. Along his *Purpose Focused Flowing Rivers, Streams & Tributaries*, the development of evaluation theories eventually points to social improvement, with various theories such as Social justice and Knowledge Generation branching out like tributaries.

Studies concerning the classification of evaluation theories aim to better comprehend the social function and evolution process of evaluation as well as to follow closely the trend of its development in order to better guide the practices. Until now, many classifications still lack a clear line of distinction. As Donaldson pointed out on the 2012 annual meeting of American Evaluation Association, these categorizations are not sufficiently accurate, especially when without support of bibliometrics.

MAPPING KNOWLEDGE DOMAIN CONSTRUCTION BASED ON CITATION & CO-AUTHORSHIP

Mapping knowledge domain is a series of graphics displaying the development, structure and connections of knowledge. It illustrates the development history, knowledge structure and frontiers, etc. of disciplines with visual images, and provide a unique prospective for portraying domanial knowledge structure. This study will first construct evaluation theory maps through citation analysis of leading researchers.

CONSTRUCTING SET OF LEADING RESEARCHERS

To construct disciplinary mapping knowledge domain through citation analysis, the first step is to select related authors and form a representative set. Currently, the most common method is to select authors according to their citation frequency, or say, to decide on a threshold of citation frequency and to choose authors that are up to such standards.

In this study, the author defines leading researchers as scholars who have made original, groundbreaking and significant contributions in this field. Combined also with the Innovation Capability Index Quantitative Monitoring System of CAS and studies of scholars like Anne E. Heberger (2010), the candidates for the set of leading researcher should contain frequently cited authors and also these researchers:

1. Researchers who hold key positions in major academic organizations, including chairpersons, vice chairpersons and secretary-generals of AEA, EES, CES and AES.
2. Recipients of important awards, which are mainly Evaluation Theory Award, Outstanding Evaluation Award, Evaluation Practice Award and Promising New Evaluator Award of the AEA Annual Meeting.
3. Chief editors and associate editors of important periodicals, including journals of AEA and CES, key journals selected by Heberger et al, and other 16 English journals found in the SSCI database when searched with keywords like "evaluation", "assessment" and "measurement".

Following such a method, we found 139 candidates for the set of leading researcher, and then further investigated their publications, including books, papers and research reports. Due to the fact that the publication and citation of books and research reports are inadequately recorded in database and that their main viewpoints are most usually covered in papers, in this study, discussion about leading researchers' publications concern their journal articles only. We chose Villanova University's library (https://library.villanova.edu) as our research database for its comprehensive collection of evaluation publications. Using the names of these researchers as keywords, we searched for their publications in the field of evaluation and found 1929 papers published on 284 periodicals.

ANALYSIS BASED ON CO-CITATION

Co-citation Analysis used to study discipline evolution and theoretical development appeared in the 1970s. In 1973, information scientist Irina Marshakova of the former Soviet Union and American scientometrics scientist Henry Small, respectively, observed this new method, co-citation analysis, as a way to measure relationships between documents. When article A and B are both cited in the bibliography of article C, they become co-cited. And if we extend Leo Egghe's criteria about co-citation analysis to authors, there can be these two criteria.

> **Criteria A**: Among a group of authors, if every one of them is co-cited with at least one other author, they form a co-cited cluster of researchers.
>
> **Criteria B**: Among a group of authors, if they are all co-cited with a given author at least once, they also form a co-cited cluster of researchers.

According to these criteria, we can model disciplinary scientific knowledge structure with co-citation relationships, and thus portray the relations between key concepts and important ideas within research fields. In other words, when two authors are cited together by a third author, they have a co-citation relationship. Authors that are frequently co-cited tend to be interrelated in terms of the concepts, theories and methods concerning their research topics. And higher co-citation frequency indicates such relationship is stronger. In order to make this co-citation based disciplinary mapping knowledge domain more reliable and accurate, this study requires rather full co-citation data. Because of limited access to this information, our data source of raw co-citation data is Microsoft Academic Search.

Then we constructed a co-citation matrix of leading researchers, and, using chi-square statistics, intra-group connection and hierarchical clustering, grouped these 50 leading researchers into 6 clusters with Pajak software (indicated by different colors in figure1).

MAPPING KNOWLEDGE DOMAIN BASED ON CO-AUTHORSHIP

It is generally assumed that scholars who often co-author publications tend to have similar concepts, theories and methods in terms of their research topics. Thus we employed Pajek, software to visualize and analyze networks, and illustrated co-authorships between scholars using Kamada-Kawai layout (figure1). In this figure, the lines connecting two scholars represent their co-authorship and thickness of the lines is proportionate to the amount of publications co-authored (the length of the lines does not represent anything).

FIND THE CORE AUTHORS IN EACH CLUSTER

Core authors in disciplinary mapping knowledge domain refer to authors of frequently-cited publications in a certain research area. These authors tend to be those who have raised new theories, methodologies and applications. As a result, they receive much attention and their core publications are frequently cited. Here, we further investigate the core scholars of each cluster in figure1 based on results gathered from previous cluster analysis of co-citations.

Then, we visualized the co-citations between scholars in figure2. Here, arrows start at the citers and point to the authors cited; and the size of arrows represent the citation frequency, which forms a mapping knowledge domain to portray the reference relationships between these scholars. We define core authors as scholars who meet these two criteria simultaneously: they should be cited by at least three other scholars from the leading researcher set; and every one of their publications in this area should be cited at least 20 times. Consequently, here are the core scholars of each cluster (as circled in figure2) in the mapping knowledge domain.

CLASSIFICATION OF LEADING RESEARCHERS' RESEARCH FRONT AND RESEARCH AREA

When mapping scientific structures, the research areas and research fronts of a discipline are usually measured with co-citation analysis based on bibliometrics. In more specific terms, researchers analyze the co-citation relationships between frequently cited publications and group them into certain clusters of papers, which are called "Research front (or RF)". Then, through further cluster analysis, these research fronts are classified again to form larger clusters of publications, which are called "Research area (or RA)".

In this research, we borrowed some ideas from the above-mentioned studies. In terms of leading evaluation researchers that this study hopes to cast light on, with co-citation analysis, we can only measure their research front but cannot further investigate their research areas, which is due to the following two reasons: on the one hand, size of the sample is rather small, which means cluster analysis is of low value; on the other hand, scholars of evaluation research tend to conduct studies in a wide variety of subjects. Thus, we need further investigation and generalization to measure their research areas. Therefore, we based this study on Marvin C. Alkin's classification (2004), and started with core authors as well as their core publications. In order to measure the research fronts and research areas of leading researchers, we analyzed their co-authorships, reference relationships and co-citations successively.

CORE AUTHORS

From visualized reference relationships presented above, we can see that J. Bradley Cousins is on the center of the core author set. Surrounding Cousins, others are located on two rings, the first ring (or say the inner ring) consists of these five scholars, Jennifer C. Greene, Michael Scriven, Leonard Bickman, Peter H. Rossi and Donald T. Campbell. And the three other authors on the second ring (also the outer ring) are as follows, Ernest R. House, Eleanor Chelimsky and Huey-Tsyh Chen. According to their research contents, we classified these nine scholars into the following three research areas of the Use Area, the Methods Area and the Valuing Area.

OTHER SCHOLARS IN THE LEADING RESEARCHER SET

As for other scholars, according to their co-authorship, reference relationship and co-citation relationship, as well as semantic analysis of their representative publications, we can determine their research areas, research fronts and main viewpoints. Also given in the areas of use, Method and valuing.

THE STRUCTURE AND CHARACTERISTICS OF THE EXTENDED THEORY TREE

In the current complex social system, evaluators contribute greatly to management and communication, playing an important role in organizational improvement. The concept of social responsibility laid a solid sociological foundation for the development of evaluation. Social needs urged evaluation theories to develop, and increasingly refined epistemology further facilitated their evolving. Now, evaluation theory, with its deep roots in social needs, social responsibility and epistemology, has made great progress in the past a few decades. Researchers are constantly reaching out into new research fronts; no matter in the use, methods or valuing area. In this research, the author, through the use of mapping knowledge domain, extended Alkin's (2004Ð2012) evaluation theory tree in these three aspects.

First of all, the extended theory tree categorizes more scholars according to their research fronts and presents them respectively, to be more exact, it can now cover 48 researchers in three research areas and 16 specific research fronts. Secondly, its branches symbolize the historical development of hot topics and related researchers. Based on its original three primary branches, the author extended the tree with results of co-citation cluster analysis and semantic analysis of frequently cited publications. As a result, on the extended theory tree, main braches represent research fronts and sub-branches stand for hot topics. On these braches writes the name of related researchers, indicating their contribution to lay the foundation. Besides, leaves on the braches stand for the representative researchers of each hot topic. With such a progressive structure, the extended theory tree can better illustrate the development of evaluation theories, especially the inheritance relationship between scholars in different research areas. Thirdly, the extended theory tree made some corrections on the classification of one or two scholars' research areas. In other words, according to the chronological order in which scholars author these theories, their citation frequency, co-citation cluster analysis and reference relationship, combined with quantitative analysis and some expert judgments, some scholars were relocated on the evaluation theory tree. For example, Stufflebeam, as can be seen in figure3, holds a more important position in the theoretical investigation of the valuing area, judging from no matter cluster analysis, reference relationship or semantic analysis of his representative publications. Moreover, this study also added to the theory tree another 21 leading researchers who have been making active contributions to evaluation theories and practice in the recent years, and omitted a handful of researchers who do not accord to the criteria of leading researchers.

CONCLUSION

This study, through the use of mapping knowledge domain, extended Alkin's evaluation theory tree, which provides valuable reference for finding out about the historical development and recent progress of evaluation theories. But there are also practical limitations in the use and extension of mapping knowledge domain. On the one hand, the clusters of leading researchers are not complete. Though standardized methods and specific indicator system have enabled us to identify most of the leading scholars, there are, unavoidably, some scholars that we have missed out. For example, Thomas Owens and Robert Wolf, researchers that are quite influential in the valuing area of evaluation studies, were not included in our analysis framework because of their inadequate amount of publications. While Laura S. Hamilton and William B. Michael, who were originally covered in our framework, did not eventually show up on the extended evaluation theory tree, because their research do not feature the common characteristics of other evaluation theory studies. On the other hand, some data concerning the publications and citation frequency of leading scholars were missing. Moreover, due to limited data, this study was unable to cover the books of leading researchers, which requires follow-up studies to explore better data retrieving methods and update the cluster of leading researchers. In this way, the evaluation theory tree will be constantly evolving and increasingly refined.

AUTHORS

LI XINGRUI
Beijing Foreign Studies University, Beijing 100089
P. R. China

LI QIANG
Corresponding author:
Institute of Policy and Management
Chinese Academy of Sciences (CAS.)
No.15 ZhongGuanCunBeiYiTiao Alley, Haidian District, Beijing 100190,
P. R. China.

Phone:+86-10-59358816, +86-18601024016
Email: lq@casipm.ac.cn

TECHNOLOGY READINESS LEVEL (TRL) MAPPING FOR SYNTHETIC BIOLOGY PRODUCTS

OLGA RADCHUK AND MARKUS SCHMIDT

INTRODUCTION

Synthetic biology aims to bring a large number of new bio-based solutions to the market, such as new pharmaceutical products, chemicals and fragrances with the help of deeply engineered microorganisms.

Synthetic biology is defined as the application of science, technology and engineering to facilitate and accelerate the design, manufacture and/or modification of genetic materials in living organisms (SCENIHR, 2014).

A small number of products and applications, manufactured with the help of synthetic biology, has already reached the market or is approaching it. As synthetic biology comprises a set of rapidly advancing technologies, the number of such products will rapidly increase in the future. However, concerns – either the real or perceived - on social and health issues, surrounding these technologies, tarnish clear understanding of the market potential and the real position of the products, harbouring such technologies. A clear and transparent categorisation of synbio products on the market is therefore needed to track the level of their technologic development and determine the exact position of each product of concern. This will not only define the measures still needed to commercialise such products, but also help to identify potential problematic societal issues. Such a map will enable a public dialogue concerning possible environmental, societal or economical issues of future application and help in assessing the urgency by which such a debate has to be carried out. In the past some applications that were very early stage created a massive media attention (e.g. the recreation of extinct animals) while others were close to market and have received hardly any focus of interest.

TECHNOLOGY READINESS LEVELS (TRL)

In order to create such categorisation, we propose to apply the concept of Technology Readiness Levels (TRLs), usually used to measure the degree to which a certain technology is ready to be applied in a real environment (Pacific Northwest National Laboratory, 2012; Belina, Giesko, Karsznia, Mazurkiewicz, 2015).

Originally developed by NASA for space applications, the TRL methodology is now increasingly adapted for use in innovative economies that rely on research and development-based technologies (Pacific Northwest National Laboratory, 2012). It provides a measurable proof of successful transition of a technology from the stage of conceptualization to the full integration into the market, subdivided by 9 steps, each of which has a number of specific maturity indicators and helps to mitigate transition risks: higher levels of technology readiness might indicate lower implementation risks.

According to the recommendations of the European Space Agency for the assessment of space applications, TRL is conducted using a set of questions that can be modified for a particular field of application (ESA TEC-SHS, 2008). They reflect „exit criteria", allowing the product to move to the next TRL. The data, necessary for the assessment, include the description (the key technology being assessed, other technologies that might interact with the key one, and their interactions if appropriate), system/mission requirements (characteristics of the operation environment and performance objectives), validated R&D results (quantitative data from testing, performance characteristics etc.) and viability (perspectives of the technology under consideration and further R&D activities, related to it: R&D risk, associated with technology and the effort, needed to bring the technology to the market).

In order to land on the market, new products must be approved by the corresponding authority, which marks the transition from the TRL 8 to the TRL 9 (the last one). The authorisation requirements depend on the nature of the product and encompass all available administrative data concerning the substance of interest including its characteristics, as well as the information, relevant for the purpose of the risk assessment and risk management (Directive 2001/83/EC, 2001; Regulation No 726, 2004; Regulation No 234, 2011).

TRL FOR SYNTHETIC BIOLOGY

Classification of the requirements for synthetic biology products for different TRLs presents a significant challenge at the moment due to the lack of successfully marketed products in the EU (most of them are only approaching the market).

In order to assist the TRL classification we propose „exit criteria" for all 9 TRL of the novel products, manufactured using synthetic biology and based on the TRLs definition, issued by the European Commission (Technology Readiness Levels, 2014).

By selecting four representative applications we show the criteria that have to be met in order to arrive at a higher TRL level (see Appendix).

The system presents an important first step in tracking the progress of such products as they reach the end user. A clearly identified position of the product along the scale of development could help to estimate the costs, necessary to bring the product or the application to the market, assess its current level of technical maturity and identify the „bottlenecks" that could potentially hinder the process of development. Such classification helps to visualize the up-to-date position of the item under consideration and estimate the risk, associated with its further development. The system allows incorporating of risk assessment and management measures on different stages of product maturity together with studies of ethical and societal impacts of the new product.

REFERENCES

Belina, B., Giesko, T., Karsznia, W., & Mazurkiewicz, A. (2015). A practical application of a method for the evaluation of implementation maturity and commercial potential in R&D projects. Problemy Eksploatacji.

Commission Regulation (EU) No 234/2011. (2011). Official Journal of the European Communities, L 64, 15-24.

Development of technology readiness level (TRL) metrics and risk measures. Pacific Northwest National Laboratory, 2012.

Directive 2001/83/EC. (2001). Official Journal of the European Union, L 311(44), 11.

ESA TEC-SHS (2008). Technology readiness levels handbook for space applications, TECSHS/5551/MG/ap, 1, Rev. 6.

Regulation (EC) No 726/2004. (2004). Official Journal of the European Union, L 136(47), 1-33.

SCENIHR (Scientific Committee on Emerging and Newly Identified Health Risks), SCCS (Scientific Committee on Consumer Safety), SCHER (Scientific Committee on Health and Environmental Risks), Synthetic biology I definition. Opinion, 25 September, 2014.

Technology Readiness Levels. (2014). Horizon 2020, work programme 2014-2015.

AUTHORS

OLGA RADCHUK
Contact author
Biofaction KG
Kundmanngasse 39/12, 1030 Vienna
Austria
mob. +436805015016

MARKUS SCHMIDT
Biofaction KG
Kundmanngasse 39/12, 1030 Vienna
Austria

APPENDIX

TECHNOLOGY READINESS LEVELS FOR SYNTHETIC BIOLOGY PRODUCTS

TRL	Definitions	Food flavours and additives	Pharmaceutical synbio-based products	Synbio-based feed compounds	Microbial fuel cells
1	Basic principles observed	A need for a new product has been identified. A scientific principle to satisfy this need has been identified or discovered.	A need for a new medicament has been identified. A scientific principle to satisfy this need has been identified or discovered.	A scientific principle to produce a new feed compound has been identified or discovered.	A scientific principle has been identified or discovered.
2	Technology concept formulated	Technical feasibility to apply the new principle in order to obtain the product has been proven. The methodology and protocols have been developed.	Technical feasibility to apply the new principle in order to obtain the product has been proven. The methodology and protocols have been developed.	Technical feasibility to apply the new principle in order to obtain the product has been proven. The methodology and protocols have been developed.	Technical feasibility to apply the new principle in order to obtain the product has been proven. The methodology and protocols have been developed.
3	Experimental proof of concept	A recombinant host able to produce the target compound has been identified or developed. Preparatory quantities of the compound have been obtained. The used methodologies enable the production as intended. Risk assessment studies have been conducted.	Hypothesis testing and initial proof of concept has been demonstrated in a number of in vitro and in vivo models. All materials meet standards appropriate for their intended use. A recombinant host able to produce the substance has been identified or engineered. The applied methodology provides intended results.	Hypothesis testing and initial proof of concept has been demonstrated in a number of in vitro models. A recombinant host able to produce the target compound has been identified or developed. Preparatory quantities of the compound have been obtained. The used methodologies enable the production as intended. Risk assessment studies have been conducted.	An appropriate microorganism has been identified. Hypothesis testing and initial proof of concept has been demonstrated. Appropriate materials and media have been identified, specialized instruments for electrochemical measurements have been identified or developed.
4	Technology validated in lab	A possibility to scale the production has been proven. Safety evaluation strategy and corresponding testing strategy have been developed and implemented.	Effectiveness of the substance has been proven. Its safety has been demonstrated in a defined laboratory or animal model.	A possibility to scale the production has been proven. Safety evaluation strategy and testing strategy have been developed and implemented.	The concept has been modeled for its technical feasibility. Sufficient power density has been reached. Feasibility to scale the system has been proven.
5	Technology validated in relevant environment	Biological and toxicological studies have been conducted; safety of the product has been proven. Exposure assessment based on normal and maximum use levels has been conducted. Experiments and/or search of relevant literature have been conducted to investigate structural or metabolic similarity to flavouring substances in existing flavouring groups.	Pre-clinical studies, including GLP animal safety and toxicity, sufficient to support further trials.	Biological and toxicological studies have been conducted; safety of the product has been proven. Recommended levels of use have been established.	The model of the system works as intended. A possibility of long-term operation of an unattended system has been proven.
6	Technology demonstrated in relevant environment	Proposed uses in food and proposed normal and maximum use levels have been established. The application dossier to obtain authorization has been prepared, including administrative data, risk assessment and management data, as well as classification and sensory profile.	Phase 1 clinical trials is completed. It supports proceeding to phase 2 clinical trials.	The application dossier to obtain authorization has been prepared, including administrative data, risk assessment and management data, as well as classification and sensory profile. A proposal for post-marketing monitoring has been developed.	A preliminary design of the system and cost estimates have been developed for the relevant environment. The model system works as intended in the operation environment.
7	System prototype demonstration in operational environment	The new product has been clearly characterized. The application dossier is submitted for authorization opinion to the European Food Safety Authority (EFSA).	Phase 2 clinical trials is completed. Phase 3 clinical trial plan is approved.	The new product has been clearly characterized. The application dossier is submitted for authorization opinion to the European Food Safety Authority (EFSA).	The system prototype works as intended or the feasibility to reach this level has been proven. The last barriers for the appropriate functioning have been identified and can be eliminated.
8	System complete and qualified	The authorization by the EU member states expert committees is given. The plan for studies on ethical and societal impacts of the product is developed.	Phase 3 clinical trial is completed. A package, its labeling and a leaflet are developed. Expert opinion is provided. Environmental risk assessment has been conducted; a risk management strategy and a post-market monitoring plan have been developed. Authorization is obtained.	The authorization by the EU member states expert committees is given. The plan for studies on ethical and societal impacts of the product is developed.	The system has been manufactured and works as intended in the operation environment.
9	Actual system proven in operational environment	The product is on the market. Post marketing studies and surveillance; studies on ethical and societal impacts of the new synbio-based product are in the process.			

NETWORKING ADDITIONALITY, BUT FOR WHOM? EVIDENCE FROM A REGIONAL INNOVATION PROGRAMME

ANNALISA CALOFFI, FEDERICA ROSSI AND MARGHERITA RUSSO

The last twenty years have witnessed the diffusion of innovation policies that attempt to foster innovation by encouraging interactions between organisations with different knowledge and competencies (Mowery, 1994; Metcalfe and Georghiou, 1998; Autio et al., 2008). In particular, many of these policy interventions have been aimed at small and medium-sized firms (SMEs), encouraging them to strengthen their skills, knowledge and creative abilities by interacting with suitable partners (Davenport et al., 1998; Bougrain and Haudeville, 2002; Narula, 2004; Caloffi et al., 2015; Rossi et al., 2016). The rationale for these policies lies in the theory of system failures (Klein Woolthuis et al., 2005; Edquist, 2011). In particular, by encouraging the diffusion of knowledge through networks, they aim to address network failures that can occur whenever the lack of linkages between agents leads to an insufficient development of complementarities, learning processes, and creation of new ideas, or when agents are trapped in relational and knowledge lock-ins (Carlsson and Jacobsson, 1997; Malerba, 1997; Nooteboom, 2000; Hekkert and Negro, 2008).

The problem of how to analyse and evaluate such policies has recently entered the agenda of both researchers and policymakers. In addition to the typical objectives of programme evaluation, it would be important to investigate whether and to what extent these interventions have taught participating firms how to collaborate with other agents, given their specific aim to encourage networking. The concept of network additionality (Clarysse, 2006; OECD, 2006) can be effectively used for the purpose. This concept comes from the idea of behavioural additionality, which refers to the ability of a policy to stimulate learning processes that result in changes in the behaviour of participating organisations during and/or after the project's implementation (Buisseret et al., 1995; Georghiou, 2002). In the case of policies that support interactions between organisations, key behavioural effects would include improvements in these organisations' ability to engage in cooperation and networking (Falk, 2007; Autio et al., 2008; Clarysse et al., 2009). Analysis of R&D collaboration policies that take the perspective of network additionality have been performed in the last years (Davenport et al., 1998; Autio et al., 2008; Nishimura and Okamuro, 2011; Wanzenböck et al., 2013; Knockaert et al., 2014). However, three aspects have so far received little or no attention. The first one is related to the fact that the existing works explore the issue of behavioral additionality without going into detail on the type of agents with which relationships occur. Instead, It can be important to understand whether

the policy encourages firms to build relationships with new organisations (i.e. organisations with which they did not have any previous relationships), new types of agents (e.g., universities), or if it fund pre-existing relationships. If the policy (also) aims at providing firms with stimuli that can help them to break their cognitive and relational lock-ins it is important to open the black box of the relationships supported by politics, and to try to understand whether this objective has been achieved. Moreover, as R&D collaboration policies often aim at promoting technology transfer processes, it is important to understand whether firms have been able to collaborate with research organisations instead that with other types of agents.

The second aspect is that studies have implicitly assumed that the effect of policy on the supported firms is homogeneous, i.e. that all firms react in the same way to public incentives. Exceptions are Clarysse et al. (2009), Wanzenböck et al. (2013) and Knockaert et al. (2014), which have taken into account firms' characteristics such as experience in R&D, technological specialisation or the participation to previous R&D collaboration policies (Wanzenböck et al., 2013). Indeed, it is known that firms possess asymmetric resources and competencies, upon which they define their idiosyncratic organizational routines, including those related to the innovative process (Dosi 1997, 1998). For instance, it is reasonable to expect that firms that already possess the knowledge and competencies that are necessary for the management of external relations and the absorption of external knowledge would benefit differently from public incentives encouraging collaborations than firms that do not have such resources. Shedding light on these effects, both theoretically and empirically, is one of the objectives of this paper.

The third aspect is a methodological one. Previous evidence on network additionality has been mainly descriptive (exceptions include Antonioli et al., 2014), while we use a propensity score matching approach (Rosenbaum and Rubin, 1983) to make inference on a set of original data that we have collected through an ad hoc survey. Our approach takes into account the heterogeneity of the effects of policies in two ways. First, we define matches between treated and controls on the basis of a wide range of variables, which are able to take into account for the different ex ante propensities of firms to cooperation. Second, we try to identify what are the effects of the policies not only on the average treated firm, but also of significant subpopulations of treated firms, such as those that, before the policy, did not perform their innovative activities in collaboration with other agents, or those that did not have absorptive capacity. In so doing, we are able to measure the 'inducement effect' (Czarnitzki and Licht, 2005; Gonzáles and Pazó, 2008) of the policy on networking.

For all relevant subpopulations of firms, we are able to specify if behavioural additionality is generated by an increase in the relationships with new organisations, with new types of organisations or with organisations the firm already collaborated with.

Given that the different types of additionality are interlinked (Clarysse et al., 2009; Roper and Hewitt-Dundas, 2014), besides network additionality we also consider the effects of the policy on firms' R&D expenditures (input additionality), as well as on firms' innovative outputs (output additionality). In addition, given that learning and changing behaviour take time, we analyse what happens to firms' networking (three years) after their participation to the collaboration policy.

We adopt this approach in the evaluation of a homogeneous set of public interventions implemented in the Italian region of Tuscany between 2002 and 2008 with ERDF funds. Such policies (4 programs divided in 9 waves, participated by 1,621 firms) funded collaborative R&D projects implemented by networks of heterogeneous agents. The policymaker's goal was to support innovation in regional small and medium-sized firms (SMEs) by stimulating them to collaborate with other regional agents (universities, service

providers or other agents) and thereby facilitate the development of interorganisational learning, synergies and complementarities (Caloffi et al., 2015; Rossi et al., 2016).

Our results show that, in general, policies seem to have had a positive effect especially with respect to relationships with universities. The funded SMEs were more likely to create relationships with new types of agents such as universities, they created more intense linkages with universities than untreated firms and, finally, they expanded the range of their collaborations with university partners. Especially this last variable can say something about the additionality of the observed policies. In fact, after taking part in the funded projects, SMEs established links with university partners that were different from those that did not participate in such projects. In other words, policies have spurred SMEs to network with universities. We do not observe similar effects on the collaboration with other types of agents. Once the policies were over, the SMEs resumed their collaboration with the other firms and providers of innovative service with which they used to collaborate before the policies.

We also found that participation in the policy persistently increases firms' absorptive capacity. Indeed, SMEs either changed their organisational structure by creating an internal R&D department, probably in order to take full advantage of new collaborations, or they performed staff training activities or both.

If we look to relevant subset of firms, we note that all the positive effects described so far are generated by the presence of firms without prior absorptive capacity. On the contrary, policies do not seem to have any effect (on average) on firms' with prior absorptive capacity.

Policy seems to have a positive effect also on firms that, before the policy, had no external relationships. Indeed, we find that they persistently modify their behaviour, and continue to create inter-organisational relationships even after the end of the policies. This behaviour occurs in relation to universities and other firms, while relationships with service providers remain unchanged. In particular, the university partners after the policy were not just those with which companies had collaborated during the development of the funded projects, but also others. The same holds true for firms other than those who were partners in the funded project.

Interestingly, participation in network projects seems to have stimulated the expansion of collaborations even for firms with prior networking behaviour. Therefore, the power of Matthew effect, such that firms with prior external relationships enjoy a cumulative advantage over firms without such relationships is stronger than other possible learning effects occurring in prior non-networked firms.

Finally, as in the overall population of firms, policy participation seems to have increased the absorptive capacity of firms without prior networking behaviour, while firms with prior networking behaviour enjoy some output additionality.

In conclusion, we observe that the policy have been successful in strengthening firms' networking ability. They have stimulated the creation of relationships with new types of agents, with which firms did not have any relationships before the policy. In addition, they have generated some inducement effects on firms that before the policy did not perform any R&D activity. Only in some cases, policies have persistently stimulated firms to produce innovative outputs.

REFERENCES

Antonioli, D., Marzucchi, A., & Montresor, S. (2014). Regional innovation policy and innovative behaviour: Looking for additional effects. European Planning Studies, 22(1), 64-83.

Autio, E., Kanninen, S., Gustafsson, R. (2008). First- and second-order additionality and learning outcomes in collaborative R&D programs, Research Policy, 37(59–76).

Bougrain, F. and Haudeville, B. (2002). Innovation, collaboration and SMEs internal research capacities. Research Policy, 31(5), 735-747.

Buisseret, T., Cameron, H., Georghiou, L., 1995. What difference does it make? Additionality in the public support of R&D in large firms. International Journal of Technology Management 10, 587–600.

Caloffi, A., Rossi, F., & Russo, M. (2015). What makes SMEs more likely to collaborate? Analysing the role of regional innovation policy. European Planning Studies, 23(7), 1245-1264.

Carlsson, B., & Jacobsson, S. (1997). In search of useful public policies—key lessons and issues for policy makers. In Technological systems and industrial dynamics (pp. 299-315). Springer US.

Clarysse, B., Wright, M., Mustar, P., 2009. Behavioural additionality of R&D subsidies: a learning perspective. Research Policy 38(10),1517–1533.

Czarnitzki, D., Licht, G., 2005. Additionality of public R&D grants in a transition economy: the case of Eastern Germany. Economics of Transition 14 (1), 101–131.

Dosi, G. (1997). Opportunities, incentives and the collective patterns of technological change. The Economic Journal, 107(444), 1530-1547.

Davenport, S. Grimes, C. Davies, J. (1998). Research collaboration and behavioural additionality: a New Zealand case study, Technology Analysis and Strategic Management 10 (1): 55–67.

Falk, R., 2007. Measuring the effects of public support schemes on firms' innovation activities—survey evidence from Austria. Research Policy 36, 665–679.

Georghiou, L. (2002). Impact and additionality of innovation policy In Boekholt, P. Innovation policy and sustainable development: can public innovation incentives make a difference? IWT Observatory on Innovation Science Technology n.40, p.57-67.

Gonzales, X., Pazo, C., 2008. Do public subsidies stimulate private R&D spending? Research Policy 37(3), 371–389.

Hekkert, M. P., & Negro, S. O. (2009). Functions of innovation systems as a framework to understand sustainable technological change: Empirical evidence for earlier claims. Technological forecasting and social change, 76(4), 584-594.

Klein Woolthuis, R., Lankhuizen M. and Gilsing, V. (2005). A system failure framework for innovation policy design, Technovation 25(6): 609-619.

Knockaert, M., Spithoven, A., & Clarysse, B. (2014). The impact of technology intermediaries on firm cognitive capacity additionality. Technological Forecasting and Social Change, 81, 376-387.

Metcalfe, J. S., & Georghiou, L. (1998). Equilibrium and Evolutionary Foundations of Technology Policy. STI Review - Special issue on New Rationale and Approaches in Technology and Innovation Policy, 22.

Mowery, D. (1994). Science and Technology Policy and Interdependent Economies, Boston: Kluwer Academic Publishers.

Narula, R. (2004). R&D collaboration by SMEs: new opportunities and limitations in the face of globalisation. Technovation, 24(2), 153-161.

Nishimura, J. Okamuro, H. (2011). Subsidy and networking: the effects of direct and indirect support programs of the cluster policy, Research Policy 40(5): 714–727.

Nooteboom, B. (2000). Learning and innovation in organizations and economies. OUP Oxford.

OECD (Ed.) (2006), Government R&D Funding and Company Behaviour— Measuring Behavioural Additionality. Paris: OECD.

Roper, S., & Hewitt-Dundas, N. (2014). The legacy of public subsidies for innovation: input, output and behavioural additionality effects (No. 0021).

Rossi, F., Caloffi, A., & Russo, M. (2016). Networked by design: Can policy requirements influence organisations' networking behaviour?. Technological Forecasting and Social Change, 105, 203-214.

Wanzenböck, I., Scherngell, T., & Fischer, M. M. (2013). How do firm characteristics affect behavioural additionalities of public R&D subsidies? Evidence for the Austrian transport sector. Technovation, 33(2), 66-77.

AUTHORS

ANNALISA CALOFFI
Corresponding author
Department of Economics and Business, University of Padova
via del Santo 33
Padova
tel. +390498274067

FEDERICA ROSSI
School of Business, Economics and Informatics, Birkbeck College, University of London

MARGHERITA RUSSO
Department of Economics, University of Modena and Reggio Emilia

KNOWLEDGE GOVERNANCE AND PLACE-BASED POLICY LEARNING: THEORETICAL REFLECTIONS

ALWIN GERRITSEN AND NICOLA FRANCESCO DOTTI

Evaluation and monitoring of Research, Technology and Innovation (RTI) policies are now well-established and recognised, however, how these practices feed the policy-making process is still an open challenge, mainly whether a place-based approach is assumed as in the case of the EU smart specialisation strategies (S3). The case of the S3 policy with its emphasis on entrepreneurial discovery processes provides the opportunity to reflect on knowledge dynamics for territorial development policies. Although the S3 policy is still too recent to discuss empirical outcomes, we aim to discuss some theoretical reflections on the emerging notion of 'knowledge governance' and its implicit, contextual and capacity to act based approach to knowledge that will be applied to various cases of place-based policy learning and knowledge governance comparing S3 approaches with other cases of RTI policy. The cases that will be reflected on are the Helsinki Smart Specialization Strategy, the Brussels Innovation Strategy, and two Food Cluster strategies in Mexico and in the Netherlands. The paper explores the place-based challenge of the intrinsic limits of local epistemic communities, the need to establish trans-territorial knowledge networks, and the activation of context-specific knowledge through entrepreneurial processes. The cases are analysed for how they address 'transdisciplinary knowledge production', 'social learning', 'self-organization', 'reflexivity', 'boundary arrangements', and 'anchoring of trans-territorial knowledge in place-based innovation policies'. This will provide insights on what such a knowledge governance and place-based policy learning approach means for evaluating and monitoring of RTI and specifically S3 policies.

INTRODUCTION

Evaluation and monitoring of Research, Technology and Innovation (RTI) policy are now well-established and recognised, however, how these practices feed the policy-making process is still an open challenge, mainly whether a place-based approach is assumed as in the case of the EU smart specialisation strategies (henceforth, "S3"). The evaluation and monitoring of RTI policy, similarly to other policies, imply the acquisition and processing of knowledge about how these policies are implemented (usually, through monitoring and evaluation) to improve policymaking (the so-called 'policy learning'), and as ultimate goal to improve economic growth, sustainable development, and social cohesion. What we do know is that 'speaking-truth-to-power' notions are too simplistic (Slembeck, 1997; Dunlop et al., 2013). The case of the S3 policy with its emphasis on entrepreneurial discovery processes provides the opportunity to reflect on knowledge dynamics for territorial development policies. In this context, our contribution aims to discuss the emerging notion of 'knowledge governance' and its implications for place-based policy learning.

KNOWLEDGE AND RTI POLICIES

The increasing demand for RTI policy evaluation fits with the current idea that we are living in an 'age of knowledge' (Dzisah and Etzkowitz, 2011), where knowledge is fundamental for territorial development (Moulaert and Sekia, 2003). This insight coincided with the rapid emergence of knowledge management practices and knowledge networks to facilitate learning and innovation within and between organizations (Easterby-Smith, 2011). However, these management approaches often have implicit assumptions about the nature of knowledge (Borraz and John, 2004; Hartlapp, 2009): knowledge is often reduced to information that can be owned, shared and traded as a commodity and can be transferred with relative ease and little cost (Adolf & Stehr, 2015). Even though codified information such as data, maps, statistics, websites and reports can in fact be easily transferred, this does not automatically imply policy learning, even more in the case of RTI policy where knowledge is the intended outcome.

Recent approaches have discussed the specificities of knowledge dynamics, specifically referring to policy-making (Hoppe, 2005) and research-policy dialogues (Dotti, 2016). A major challenge for the discussions is to integrate the explicit scientific knowledge with policy related knowledge which is of 'a capacity to act' type. Knowledge as capacity to act involves skills, expertise, a model to elaborate information as well as 'tacit knowledge' (Polanyi, 1967) that is known for being context-specific, and then difficult to share with 'external' individuals. Tacit knowledge therefore has characteristics of public or collective goods because ownership cannot easily be attributed, is non-rival and hardly excludable (Adolf and Stehr, 2014). This knowledge benefits from the so-called 'comedy of commons': the more knowledge is used and shared, the more it 'increases' (Foray, 2004).

For RTI policies such as the S3, both tacit and explicit knowledge are needed to improve policy-making and implementation dynamics. Yet, these dynamics are highly complex. The main reason being that knowledge is also related to power (Foucault, 1966) and normative agendas (Haas, 1992). Following Kuhn's perspective (Kuhn, 1962), knowledge from other epistemic communities will be seen as of limited value by actors from such a community and therefore will not be considered for decision making (Hall, 1993) or its adherents will be rejected or even punished. The presence of epistemic communities is associated to 'policy paradigms' facilitating circulation of knowledge within communities, but limiting the absorptive capacity from external sources. This (partially) explains certain inertia in policy-making. Dissemination and usage of knowledge is therefore often difficult or even impossible and statements that knowledge should be useful to society, policy or economy, fail to understand the complexities of handling knowledge.

This is especially relevant to S3 and other RTI policies because here knowledge is the main policy outcome, and then potentially creating a circular and iterative dynamics with its own policy-making process. In S3, policymaking also aims to organise learning between actors with its focus on entrepreneurial discovery and by the strategic choices regarding the regional economic path and its innovation strategy (Foray et al., 2009). Monitoring and evaluation are used to optimize these activities (Joint Research Centre, 2015): the focus on data, statistics, mapping and foresight analysis provides explicit knowledge. Yet, the S3 policy involves more than monitoring and evaluation: the emphasis on entrepreneurial discovery is instrumental to promote policy learning, mainly in lagging regions. In this perspective, the S3 policy provides a unique opportunity to use RTI policy for policy learning due to the cumulative nature of knowledge (Antonelli and Quéré, 2002). Although the S3 policy is still too recent to discuss empirical outcomes, we aim to discuss some theoretical reflections on the emerging notion of 'knowledge governance' and illustrate them with reflecting on four cases.

KNOWLEDGE GOVERNANCE: AN EMERGING PERSPECTIVE FOR POLICY LEARNING

In this paper, we elaborate on what an implicit, contextual and capacity to act based approach to knowledge can add to the theoretical debate on RTI policy, specifically referring to the S3 policy. For this we introduce the emerging concept of 'knowledge governance', which differs fundamentally from knowledge management: while the latter focuses on the management of knowledge activities, the former includes the engagement of actors in innovation questioning existing policies, and ultimately enabling policy learning (Gerritsen et al., 2013). Knowledge governance can be defined as '*purposefully organizing the development of knowledge in order to deal with societal problems. Knowledge governance is aimed at creating new insights, and innovative solutions which tempt actors to leave traditional insights and practices and get away from inert interaction patterns, stalemate negotiations, and interest conflicts. Knowledge governance is also used to raise awareness and deliver suggestions that give actors a perspective on purposeful action'* (van Buuren and Eshuis, 2010, p. 284).

In this perspective, knowledge is not limited to research and information, but in principle includes all types of knowledge and actors such as public administrations, experts and other stakeholders. Knowledge governance has been used as 'the governing of knowledge' (Foss, 2011; Van Kerkhoff, 2013), but also as a distinct 'mode of governance' (Gerritsen et al., 2013; van Buuren and Eshuis, 2010) in which the governing of knowledge has the purpose of learning to enable policy change (Capano and Howlett, 2009). Knowledge governance is an addition to the more widely used modes of governance (Hooghe and Marks, 2003), such as hierarchic, market and network governance.

Other types of governance have been suggested in literature, most notably 'reflexive governance' (VoÐ & Kemp, 2006; VoÐ & Bornemann, 2011) is relevant to knowledge governance. Although commonalities exist between 'knowledge governance' and 'reflexive governance', the latter focuses primarily on feedback loops of policy interventions to policies, but includes various modes of governance, whereas the former is a distinct mode of governance, that aims for policy, societal, and business change by developing new knowledge. Knowledge governance introduces a perspective that can be used as a focused analytical lens to monitor and evaluate RTI policies as the S3, in particular to reflect on how knowledge and learning were handled and what its impacts were. Knowledge governance interventions can be characterised by the principles of 'transdisciplinary knowledge production', 'social learning', 'self-organization', 'reflexivity' and 'boundary arrangements' (Gerritsen et al., 2013). These characteristics do not constitute a 'blue print'. In each knowledge governance practice these characteristics will differ in its details because of context and because cross-overs with other modes of governance are likely to occur.

A PLACE-BASED PERSPECTIVE ON KNOWLEDGE GOVERNANCE

Knowledge governance can be applied to any kind of knowledge for policy learning process, including RTI policy following an S3 approach. Nevertheless, the S3 policy requires the introduction of a place-based perspective imposing the following challenges, often neglected in the literature.

1. The intrinsic limits of local epistemic communities. While the local scale facilitates the establishment of networks, the exchange of knowledge, and trust building, it can also determine lock-in, rejection of deviant (i.e. innovative) approaches, and long-term mistrust in case of multiple failures. This is especially challenging for local communities in peripheral areas (Lagendijk and Lorentzen, 2007) and, even more, when complex and specific expertise is needed.
2. The need to establish trans-territorial knowledge networks (Bathelt et al, 2004) due to the different scale of policy challenges, and the need to access knowledge sources not available locally. In the case of RTI, the importance of recombining existing knowledge across territories is a strategic capacity to pursue market-oriented innovation, yet it requires resources to scan for knowledge available elsewhere.
3. The context-specific knowledge can be activated through entrepreneurial processes of discovery such as S3 strategies with the important side-effect of empowering local epistemic communities. Policy learning can be strengthened when an exogenous actor as the EU Commission intervenes (e.g. with the Horizon2020 program or the ERDF technical assistance), when this is anchored in the place-based entrepreneurial practices (Crevoisier and Jeannerat, 2009).

Therefore, for RTI and S3 evaluation purposes knowledge governance needs to include the *anchoring of trans-territorial knowledge in place-based RTI policies.*

CASE REFLECTIONS

The Knowledge Governance framework will be used to reflect on four cases of policy learning and knowledge governance comparing S3 approaches with other cases of RTI policies. The cases that have been studied are the Helsinki Smart Specialization Strategy (Nissinen, in prep.), the Brussels Innovation Strategy, and Food Cluster strategy formulation and implementation in Mexico and in the Netherlands (Gerritsen, Stuiver & Termeer, in prep.; Gerritsen et al., 2015). Two of the cases are concerned with the development and implementation of regional smart specialisation strategies, in particular in Finland and Belgium. The other two cases are examples of strategy formulation and implementation in food cluster development. Although the objective of this paper is not to conclude whether the cases are good or bad examples of knowledge governance for place based policy learning, they were studied and compared with the help of the knowledge governance characteristics. The following issues have been identified by using this place-based knowledge governance approach to evaluating RTI policies:

- All cases showed difficulties in implementing a transdisciplinary approach, based on real-world problems and with the inclusion of knowledge of entrepreneurs and other stakeholders. Social learning was organised in all cases, but because of lacking transdisciplinarity this was only to a limited extent connected to the real world problems associated with smart specialization and cluster development. In two cases, transdisciplinarity in knowledge processes was established to a limited extent, but this did not dominate strategy formulation and implementation. The usage of scientific and policy concepts by key actors and attempt to keep control of the process by existing actor networks hindered this. Establishing transdisciplinarity therefore was a bottleneck in the studied cases and that is particularly relevant because S3 - and more specifically entrepreneurial discovery - is very much about transdisciplinarity.
- In all cases the RTI activities were initiated by public authorities or sector organisations, and only afterwards opened up bottom-up participation, i.e. stakeholders' participation was promoted and driven by policymakers. The initiators controlled the knowledge processes and also limited their extent. In the Finnish case corporatist governance traditions with closed already existing knowledge networks prevented the entrepreneurial discovery process to take off, in the Mexican and Dutch cases, process facilitation technics enabled this, but within set boundaries limiting the reflexive characteristics of the knowledge governance process.
- Managing the boundary with actors outside of the learning setting is very much a challenge to all knowledge and innovation processes. All cases paid attention to this, but they all struggled with this. In the Finnish case, the boundary between local companies and public actors proved too hard to cross, in the Dutch case the project infrastructure development branch of Greenport Venlo made their own choices and the knowledge generated in the initiative was not really implemented by them. In the Brussels and Mexican cases the different actor-types primarily remained in their own domains and the process managers and facilities struggled to enable synergies to emerge.
- The anchoring of trans-territorial knowledge was mainly an issue in the Mexican case and, partly in the Finnish and Dutch cases, in the sense that S3 concepts from outside of the territory were introduced in the knowledge governance process. In the Mexican case a team of Dutch scholars and business people was hired to assist them in redeveloping the food cluster. The Brussels case focussed on Brussels knowledge only, although many international knowledge and education institutes are located in Brussels, because of the presence of the European institutions.

CONCLUSIONS

S3 is in itself a form of multi actor policy learning in which innovations are developed and boundaries between actors, narratives and ways of working need to be crossed. S3 is aimed at reformulating policies by a process of entrepreneurial discovery in in that sense is a clear example of knowledge governance and also has its broad approach to knowledge as capacity to act, including as policy learning by policy actors. This clearly includes other knowledge that what can be captured in data, maps, etc. The knowledge is very much part of the place-based policy processes, the relations between actors, the relations with outside actors and networks, the selected valorisation options, etc. Especially entrepreneurial discovery has these characteristics. Entrepreneurial discovery is more than asking from entrepreneurs what should be the focus of RTI policies. These ideas have to become part of already established governance and knowledge networks that mostly are dominated by public agents and experts and therefore using entrepreneurial discovery for prioritising also involves incorporating their knowledge and possibly to the extent of other knowledge and positions. As we have shown for our cases, this is a real challenge to implement, and the selected characteristics of knowledge governance processes were useful in identifying the barriers that need to be levelled.

The results of our research imply that evaluating S3 policies is not a neutral activity, but a form of knowledge governance that has the potential to change S3 and other RTI policies and implementation. This policy learning requires a contextual and capacity to act based approach to knowledge and learning, surpassing notions as 'speaking truth to power' (Wildavsky, 1979) or evidence based policy making.

REFERENCES

Adolf, M., Stehr, N., 2014. Knowledge. Routledge.

Antonelli, C., Quéré, M., 2002. The Governance of Interactive Learning within Innovation Systems. Urban Stud 39, 1051–1063. doi:10.1080/00420980220128453

Bathelt, H., Malmberg, A. & Maskell, P. (2004) Clusters and knowledge: Local buzz, global pipelines and the process of knowledge creation, Progress in Human Geography, 28(1), pp. 31–56.

Borraz, O., John, P., 2004. The Transformation of Urban Political Leadership in Western Europe. International Journal of Urban and Regional Research 28, 107–120. doi:10.1111/j.0309-1317.2004.00505.x

Capano, G., Howlett, M., 2009. Introduction: The Determinants of Policy Change: Advancing the Debate. Journal of Comparative Policy Analysis: Research and Practice 11, 1–5. doi:10.1080/13876980802648227

Crevoisier, O., Jeannerat, H., 2009. Territorial Knowledge Dynamics: From the Proximity Paradigm to Multi-location Milieus. European Planning Studies 17, 1223–1241. doi:10.1080/09654310902978231

Dotti, N.F. (2016). A research-policy dialogue for the EU Cohesion Policy. In: Dotti, N.F., (ed.), 2016. Learning from implementation and evaluation of the EU Cohesion Policy: Lessons from a research-policy dialogue. RSA Research Network on Cohesion Policy, Brussels, 26-42.

Dunlop, C.A., Raedelli, C.M. (2013). Systematising Policy Learning: From Monolith to Dimensions. Political Studies 61: 599-619. doi: 10.1111/j.1467-9248.2012.00982.x

Dzisah, J., Etzkowitz, H., 2011. The Age of Knowledge: The Dynamics of Universities, Knowledge & Society. BRILL.
Foray, D., 2004. Economics of knowledge. MIT Press, Cambridge, Mass.

Foray, D, David, P.A., Hall, B., 2009. Smart Specialization - the Concept. Knowledge Economists Policy Brief 9, 25-29, Knowledge for Growth Expert Group.

Foss, N.J., 2011. Knowledge Governance: Meaning, Nature, Origins, and Implications (SSRN Scholarly Paper No. ID 1969400). Social Science Research Network, Rochester, NY.

Foucault, M., 1966. Les mots et les choses: une archéologie des sciences humaines: une archéologie des sciences humaines. Gallimard.

Gerritsen, A.L. Stuiver, M. & Termeer, C.J.A.M. (in prep.). Organizing and enabling knowledge trajectories in sustainable development. In: Dotti, N.F. (Ed). Knowledge, policymaking and learning for European Cities and Regions. Edward Elger Publishers. In prep.

Gerritsen, A.L., Kranendonk, R.P., Cofino, M., Lagendijk, A., Cluster management: a knowledge centred and collaborative perspective. Paper for the TCI Global Conference 2015, 3 - 6 November, 2015, Daegu, South Korea.

Gerritsen, A.L., Stuiver, M., Termeer, C.J.A.M., 2013. Knowledge governance: An exploration of principles, impact, and barriers. Science and Public Policy sct012. doi:10.1093/scipol/sct012

Haas, P.M., 1992. Introduction: epistemic communities and international policy coordination. International Organization 46, 1–35. doi:10.1017/S0020818300001442

Hartlapp, M., 2009. Learning about policy learning. Reflections on the European Employment Strategy. EIoP 13. doi:10.1695/2009007

Hooghe, L., Marks, G., 2003. Unraveling the Central State, but How? Types of Multi-level Governance. American Political Science Review 97, 233–243.

Hoppe, R., 2005. Rethinking the science-policy nexus: from knowledge utilization and science technology studies to types of boundary arrangements. Poiesis Prax 3, 199–215. doi:10.1007/s10202-005-0074-0

Joint Research Centre, I. for P.T.S., 2015. (No. 13/2015), S3 Policy Brief Series. EU Commission, Sevilla.

Kuhn, T.S., 1962. The structure of scientific revolutions. University of Chicago Press, Chicago.

Lagendijk, A., Lorentzen, A., 2007. Proximity, Knowledge and Innovation in Peripheral Regions. On the Intersection between Geographical and Organizational Proximity. European Planning Studies 15, 457–466. doi:10.1080/09654310601133260

Moulaert, F., Sekia, F., 2003. Territorial Innovation Models: A Critical Survey. Regional Studies 37, 289–302. doi:10.1080/0034340032000065442

Nissinen, M. (in prep.) From words to deeds in Finland. In: Dotti, N.F. (Ed). Knowledge, policymaking and learning for European Cities and Regions. Edward Elger Publishers. In prep.

Polanyi, M., 1967. The tacit dimension. Routledge & Kegan Paul, London.

Slembeck, T. (1997). The Formation of Economic Policy: A Cognitive-Evolutionary Approach to Policy-Making. Constitutional Political Economy 8: 225-254.

van Buuren, M.W., Eshuis, J., 2010. Knowledge governance: complementing hierarchies, networks and markets?, in: 't Veldt, R.J. (Ed.), Knowledge Democracy - Consequences for Science, Politics and Media. Heidelberg, pp. 283–297.

Van Kerkhoff, L., 2013. Knowledge Governance for Sustainable Development: A Review. Challenges in Sustainability 1, 82–93.

Voß, J.P., Kemp, R., 2006. Sustainability and reflexive governance: introduction, in: Bauknecht, D., Kemp, R. (Eds.), Reflexive Governance for Sustainable Development. Edward Elgar Publishing, Cheltenham / Northampton, MA, pp. 3–28.

VoÐ, J.P. & Bornemann, B., 2011. The politics of reflexive governance: challenges for designing adaptive management and transition management. In: Ecology & Society 16(2): 9.

Wildavsky, A.B., 1979. Speaking truth to power: the art and craft of policy analysis, Little, Brown. ed. Transaction Publishers, Boston, Mass.

AUTHORS

ALWIN GERRITSEN
Corresponding author
Wageningen Environmental Research, Wageningen University and Research
Droevendaalsesteeg 3, 6708PB Wageningen,
The Netherlands
T. +31 317 481926
E. alwin.gerritsen@wur.nl

NICOLA FRANCESCO DOTTI
Institut de sciences politiques Louvain-Europe (ISPOLE), Université Catholique de Louvain (UCL), Place de l'Université 1
1348, Louvain-la-Neuve
Belgium
E. nicola.dotti@uclouvain.be

International Centre for Innovation, Technology and Education Studies (iCite), Solvay Brussels School, Université Libre de Bruxelles (ULB),
Avenue F.D. Roosevelt 42
1050, Brussels
Belgium
E. nicola.dotti@ulb.ac.be

EVALUATING CITIZEN SCIENCE AT PROGRESS AND IMPACT LEVEL: WHAT'S THE VALUE FOR RESEARCH FUNDING POLICIES?

BARBARA KIESLINGER, TERESA SCHÄFER, GERTRAUD LEIMÜLLER, FLORIAN HEIGL AND DANIEL DÖRLER

Citizen science is growing in popularity across Europe. In Austria alone, the citizen science portal "Österreich forscht"[1], which clusters citizen science projects across all domains, is experiencing a considerable increase in participation, with a constantly growing number of projects (from 9 in February 2015 to 30 in April 2016). While ecology still dominates the topics of citizen science projects, new emergent areas like medical research and social sciences can be observed. On "Österreich forscht" we thus can find a project where scientists and citizens collaborate in finding out reasons for beehive death side by side with a project that aims at defining new research questions related to mental health.

The effort of opening science, the gaining interest of people to engage in science and its embedding in the wider trend of conducting responsible research and innovation (RRI) as well as the ease of access to latest ICT (Information and Communication Technology) all contribute to a shift towards participatory science and a growing importance of open innovation. This also resonates with policy makers and results in new and upcoming funding mechanisms, such as Young Citizen Science[2] and TOP Citizen Science[3] in Austria or OPAL[4] in the UK.

Citizen science is a highly dynamic approach where constantly new forms of collaboration between science and society evolve. The sheer diversity of citizen science projects makes it important to take care in drawing comparisons or using individual projects as examples or proxies for the overall phenomenon. The type of scientific work and geographic scale of participation strongly shapes project goals as well as strategies to meet these goals. There are also differences to observe between online and offline projects with regard to process and impact.

This diversity puts a challenge to ways of evaluating citizen science. The main European stakeholders in citizen science clearly identified the need for evaluation of citizen science projects and proposals (as can be found in the White Paper on Citizen Science for Europe (Serrano Sanz et al. 2014)). Evaluation helps to proof the impact of participatory research methods for science, individual participants and socio-ecological systems, and it supports a wider use and acceptance amongst all stakeholders. A comprehensive but flexible evaluation framework is also required for research funding to decide on the funding of future citizen science projects.

The public consultancy leading to the White Paper on Citizen Science for Europe (ibid) also showed that there were quite controversial opinions on how to best evaluate projects and proposals. Evaluation concepts for citizen science need to be expanded to capture the added value generated by an open, participatory research process and need to equally support different types of citizen science projects (Schäfer & Kieslinger 2016). Those projects driven more by professional scientists, where citizens mainly contribute to data collection or data processing, need to be treated equally to those projects, where research questions are elaborated in a collaborative process between professional and amateur scientists, or grassroots initiatives, where citizens or civic organisations are driving the research process themselves.

STATE-OF-THE ART OF EVALUATING CITIZEN SCIENCE

There are currently no commonly established indicators to evaluate citizen science except the first attempts from Haywood and Besley (2014) to integrate indicators for evaluating projects in participatory science. Individual citizen science projects still struggle to define the most appropriate road towards collecting evidence of their impact. Existing studies tend to focus on the scientific impact of citizens science projects and on the wider outcomes for individual participants. Especially learning gains on individual level are elaborated in more detail and are reported to be occurring across various project types (e.g. Holocher & Kieslinger 2014, Wiggins & Crowston 2015, Ziegler, Pettibone et al. 2015).

Bonney et al. (2009) recommend a two-way evaluation of scientific outcomes of the projects (e.g. number of papers) and of the learning effects for the participants (e.g. improved skills). The evaluation criteria suggested by Phillips et al. (2014) assess individual learning outcomes like any gains in scientific knowledge or skills, as well as a wider per-

1 http://www.citizen-science.at
2 http://www.youngscience.at/young_citizen_science/forschung_zum_mitmachen/
3 https://www.zentrumfuercitizenscience.at/en/top-citizen-science.html
4 http://www.opalexplorenature.org/aboutopal

sonal impact on a participant's behaviour, interests in science, motivation and self-efficacy to participate in science. Aspects addressed under the heading of behavioural change, such as taking stewardship and civic action, which all point towards social implications, are also covered by other authors (Crall, 2011).

Shirk et al. (2012) recommend a more holistic approach to project evaluation, considering the impact on the scientific knowledge gain, the individual development, as well as broader socio-ecological impact and thus consider societal, ecological, economical and political influence factors during the evaluation process.

In a similar vain, Jordan et al. (2012) promote evaluation that goes beyond learning outcomes and suggest looking moreover into programmatic and community level outcomes. Their suggestion for a more comprehensive approach to evaluation stresses the potential impact of citizen science on social capital, community capacity, economic impact and trust between scientists, managers and the public.

Taking a closer look at how evaluation of citizen science projects is conducted currently, it can be seen that data tends to be collected via surveys, interviews and the analysis of personal communication with the participants (Gommerman & Monroe, 2012). Experts recommend not to apply the same set of criteria equally across single projects but rather to define an appropriate evaluation strategy according to the goals set by each project, aligning measurable indicators (Jordan et al. 2012; Phillips et al. 2014; Tweddle et al. 2012). Wright (2011) emphasises the role of evaluation in adaptive project management. Continuously sharing experiences and lessons learned across the various stakeholders supports the social learning process and contributes to an iterative improvement of citizen science projects and programmes.

In general, comprehensive evaluation frameworks that would allow for comparability across projects and programmes while offering flexibility for adaptation are missing (Bonney et al. 2009, Bonney et al. 2014, Crall et al. 2012). Jordan et al. (2015) critically mention a lack of criteria and methods to assess the democratisation of science and its benefits for society, making it difficult to show the direct and indirect impact of citizen science on society and the environment.

THE APPROACH TOWARDS AN EVALUATION FRAMEWORK FOR AUSTRIAN FUNDING AGENCIES

In Austria the Federal Ministry of Science, Research and Economy started the elaboration of an evaluation framework to be used for future funding programmes related to RRI, citizen science and open innovation in science. With this aim three concepts were developed independently by three author teams, each having a slightly different focus and all being authors of this abstract (Kieslinger, Schäfer & Fabian 2015, Heigl & Dörler 2015, Leimüller & Swanson 2015). While one author group was focusing on covering social and learning perspectives, the second group was concentrating on the scientific process and ecological perspectives, and finally, the third group put their emphasis on the open innovation perspective in science.

The outcomes were three partly overlapping, partly diverging sets of evaluation criteria for citizen science and open innovation projects in science that covered the process and strategy as well as the outcomes and impact level. In a next step an approximation across the three different concepts in terms of comparison and mapping was initiated with the final aim to provide and integrated evaluation framework. Such a framework is defining criteria on scientific, individual actor- and project-related as well as socio-ecological dimensions. These criteria are looking into the whole process of setting up and running a project as well as assessing its scientific and societal outcomes and potential impact across the three concepts. The original concepts complement each other strongly in the area of societal benefits and ethical dimensions, but diverge on how much additional scientific value should be produced by open and participatory forms of science and on the evaluation of the appropriateness of strategies and procedures for openness and participation.

Challenges during this process were firstly to develop generic evaluation criteria that are able to cover a high diversity of citizen science projects in regard to disciplines, contexts, participatory methods and scientific goals. Secondly, scientific projects should not be overloaded with a broad spectrum of requirements. The individual strength of scientific projects (e.g. showing a clear focus and following an internal logic) cannot be sacrificed to the aspiration that projects need to be good in complying with all criteria at the same level (e.g. demonstrating equally high scientific, societal, economical and ecological value), which could destroy specific strengths. Funding agencies that focus on supporting a diversity of projects with individual strengths need evaluation criteria which are at the same time flexible and stripped-down, focusing on the core elements of citizen science and open innovation projects.

EVALUATION FRAMEWORK APPLICABILITY AND USEFULNESS FOR POLICY AND PROJECTS

In our contribution we will discuss the three concepts in more detail, show the commonalities and differences between the approaches and discuss the challenges that come with applying an integrated evaluation framework for the evaluation of citizen science and open innovation projects and proposals. We will also provide insights into how to advance from a comprehensive set of evaluation criteria to developing an easy-to-use instrument for the self-assessment of citizen science projects and project ideas. The main aim of such a self-assessment instrument is to support different types of open participatory scientific projects in reflecting about their individual strengths and shortcomings on a scientific, individual actor- and project-related level as well as on a socio-ecological and potential economic level.

REFERENCES

Bonney, R., Ballard, H., Jordan, R., McCallie, E., Phillips, T., Shirk, J., & Wilderman, C. C. (2009): "Public participation in scientific research: Defining the field and assessing its potential for informal science education". A CAISE inquiry group report. Washington, D.C.: Center for Advancement of Informal Science Education (CAISE).

Bonney, R., Shirk, J.L., Phillips, T. B., Wiggins, A., Ballard, H. L. (2014): "Next steps for Citizen Science". Science. Vol. 343: 1436-1437.

Crall, A. W., Jordan, R., Holfelder, K., Newman, G. J., Graham, J., Waller, D. M. (2012): "The impacts of an invasive species citizen science training program on participant attitudes, behavior, and science literacy". Public Understanding of Science. 22:745-764.

Gommerman, L., & Monroe, M. C. (2012): "Lessons learned from evaluations of citizen science programs" FOR291, School of Forest Resources and Conservation Department, UF/IFAS Extension. (May), 1–5. Retrieved on 25 April 2016 from https://edis.ifas.ufl.edu/pdffiles/FR/FR35900.pdf

Haywood, B. K., and Besley, J. C. (2014). Education, outreach, and inclusive engagement: Towards integrated indicators of successful program outcomes in participatory science. Public Underst. Sci. 23, 92–106.

Heigl, F., Dörler, D. (2015): "Qualitätskriterienkatalog für Citizen Science Projekte in den Naturwissenschaften". Studie im Auftrag des Bundesministeriums für Wissenschaft, Forschung und Wirtschaft (bmwfw).

Jordan, R., Crall, A., Gray, S., Phillips, T., Mellor, D. (2015): Citizen Science as a Distinct Field of Inquiry. BioScience. Vol. 65, No 2: 208–211. doi:10.1093/biosci/biu217

Jordan, R., Ballard, H., Phillips, T. B., (2012): "Key issues and new approaches for evaluating citizen-science learning outcomes". Frontiers in Ecology and the Environment, 10(6): 307–309, doi:10.1890/110280.

Kieslinger, B., Schäfer, T., Fabian, C.M., (2015): "Kriterienkatalog zur Bewertung von Citizen Science Projekten und Projektanträgen". Studie im Auftrag des Bundesministeriums für Wissenschaft, Forschung und Wirtschaft (bmwfw).

Leimüller, G., Swanson, A. (2015): "Entwicklung eines Kriterienrasters für Citizen Science und Open Innovation". Studie im Auftrag des Bundesministeriums für Wissenschaft, Forschung und Wirtschaft (bmwfw).

Phillips, T.B., Ferguson, M., Minarchek, M., Porticella, N., Bonney, R. (2014): "Users Guide for Evaluating Learning Outcomes in Citizen Science". Ithaca, NY: Cornell Lab of Orinthology.

Schäfer, T., Kieslinger, B. (2016): "Supporting emerging forms of citizen science: a plea for diversity, creativity and social innovation". Journal of Science Communication (JCOM) 15 (02), Y02.

Serrano Sanz, F., Holocher-Ertl, T., Kieslinger, B., García, F.S. and Cândida G. Silva (2014): "White Paper on Citizen Science in Europe". Socientize Consortium.

Shirk, J. L., Ballard, H. L., Wilderman, C. C., Phillips, T., Wiggins, A., Jordan, R., McCallie, E., Minarchek, M., Lewenstein, B. V., Krasny, M. E., Bonney, R. (2012): "Public participation in scientific research: a framework for deliberate design". Ecology and Society 17(2): 29. http://dx.doi.org/10.5751/ES-04705-170229

Stilgoe, J., Owen, R., Macnaghten, P. (2013): "Developing a framework for responsible innovation". Research Policy 42(9), 1568–1580.

Tweddle, J.C., Robinson, L.D., Pocock, M.J.O., Roy, H.E., (2012): "Guide to Citizen Science: developing, implementing and evaluating citizen science to study biodiversity and the environment in the UK". Natural History Museum and NERC Centre for Ecology & Hydrology for UK-EOF.

Wiggins A., Crowston, K., (2015): "Surveying the citizen science landscape". First Monday, Volume 20, Number 1 - 5 January 2015. Retrieved on 25 April 2016 from http://firstmonday.org/ojs/index.php/fm/article/view/5520/4194 doi: http://dx.doi.org/10.5210/fm.v20i1.5520

Wright, D. (2011): "Evaluating a citizen science research programme: Understanding the people who make it possible", MSc thesis, University of Cape Town, 1–124.

Ziegler, D., Pettibone, L., Rettberg, W., Feldmann, R., Brand, M., Schumann, A., Kiefer, S. (2015): "Potential für lebenslanges Lernen". Weiterbildung 2, 18-21.

AUTHORS

BARBARA KIESLINGER
Centre for Social Innovation
Linke Wienzeile 246, 1150 Vienna
Austria

TERESA SCHÄFER
Centre for Social Innovation
Linke Wienzeile 246, 1150 Vienna
Austria

GERTRAUD LEIMÜLLER
Winnovation
Karl-Schweighofer-Gasse 12, 1070, Vienna
Austria

FLORIAN HEIGL
University of Natural Resources and Life Sciences
Gregor-Mendel-Straße 33, 1180 Vienna
Austria

DANIEL DÖRLER
University of Natural Resources and Life Sciences
Gregor-Mendel-Straße 33, 1180 Vienna
Austria

SOCIETAL VS. ACADEMIC IMPACT?

A CRITICAL DISCUSSION BASED ON THE EXPERIENCES FROM EVALUATIONS OF THE “SPARKLING SCIENCE” PROGRAMME AND THE “YOUNG SCIENCE” PROJECT AND OTHER CITIZEN SCIENCE PROJECTS

ISABELLA WAGNER, CAROLINE MANAHL AND SUSANNE DOBNER

Our contribution discusses the potentials and possibilities of including societal outcomes beyond mere academic impact dimensions of those RTI programmes that specifically include societal stakeholder groups other than the scientific community, for example projects or programmes including students or pupils to the different phases of a research project.

Based on experiences from Austrian cases, like the “Sparkling Science” programme[1] or the “Young Science” project[2], as well as other (European) citizen science projects it was examined how traditionally measures of academic impact do not fully reflect their potential immediate societal impact. By presenting and discussing the methodologies and results of the evaluation studies of these programmes and projects recently conducted by ZSI, reflection on what potential societal impact was acknowledgeable with the methodologies and approaches utilised was offered.

In the Sparkling Science evaluation study (Manahl et al., 2016), the contractor particularly asked for the evaluation of the academic impact of their programme. In detail, the impact of the programme on the scientific output, on the researchers' career development, on the development of new research questions and follow-up projects and on the development of science communication competences. Methodologically, an experimental design with a control group study was used to compare the impact of the scientific publications produced within and outside the programme, an online survey, qualitative interviews and focus groups were applied.

The Young Science evaluation study (Manahl et al., 2015) focused on the networking work of the Young Science Initiative situated at the interface between school and science. More specifically it looked at the inclusion of the different stakeholders; the variety and adoption of thematic focus areas; and the user orientation in the offered services. The methodology set was comprised by a choice of qualitative measures, mainly workshops and qualitative interviews, as well as the analysis of quantitative data, utilising event data, website statistics, etc.

In both studies, as well as in earlier comparable evaluation studies (cf. Birke 2014 and 2013), it became evident that the value of including students in scientific research projects was manifested in various dimensions, but not exclusively in mere scientific output. Some examples for societal and social benefits range from a changing public attitude towards science and an improved understanding of scientific concepts, raised awareness of pupils and students about scientific careers, over knowledge transfer, to a higher intrinsic motivation on the scientists' side. There are various motives for conducting a citizen science project and the emerging benefits might differ between the different scientific fields and application contexts.

In the framework of the two evaluation studies in focus, the scientists and researchers interviewed underlined the lack of acknowledgement of “social impact” or science communication measures beyond scientific publications applied in collaboration in the work with students (e.g.: Manahl 2015, 3). While assessing the value of including citizens or particular students in scientific research itself has been asked for in evaluation studies before, with the increasing importance of citizen science in science policy, assessing long-term societal impact is being considered as gaining importance as well.

Contributing to the development of indicators in the Austrian context, Kieslinger et al. identified and elaborated on three dimensions for citizen science programmes' impact: 1) academic dimension, 2) citizen dimension and 3) the socio-ecological system's dimension. With the broader societal context in focus, this third dimension of socio-ecological impact dimensions include societal effects like social capital or political participation, ecological effects and even innovation-related impacts like cross-fertilisation with the development of new (communication) technologies, sustainability and socially innovative practices, emergence of economic potential and the development of market chances (Kieslinger et al. 2015).

The observations of the two evaluation studies were contextualised and discussed behind the various roles science and academia are playing within societies (cf. “third mission” of universities). While the science-society entanglement in citizen-science project is well intended and both, impacts to academia as well as societal impacts are measurable, the implementation of science in and for a wider society inevitably creates tensions with the “first two” traditional missions of universities (cf. e.g. Laredo, 2007).

Apart from that, there are various inherent problems in measuring societal impact of research and innovation that have been discussed earlier and go beyond the mere feasibility and resource problems due to the need of a long-term approach as well as broad methodological set of tools for a broad societal coverage in general. Such problems in soci-

1 Programme: http://sparklingscience.at/ | Evaluation study: https://www.zsi.at/en/object/project/3816
2 Project: http://www.youngscience.at/ | Evaluation study: https://www.zsi.at/de/object/project/3401

etal impact measurement reach from causality problems, since it is not always evident which effects are derived by which cause(s); attribution problems emerging due to the complexity and diffuse nature of societal impacts; problems of internationality behind the background of a globalised science production and therefore horizontal and geographic attribution problems; as well as problems of timing or time scale coordinating evaluation and policy cycles. If the impact is measured too early, the results might be premature, leading to policy recommendations that are unable to appreciate and support long-term benefits; while on the other hand, attribution and causality vagueness might increase with the time elapsed (cf. Bornmann 2012). For measuring citizen-science programmes effectiveness and general societal impacts like effects the public understanding of science it has been proposed to expand the analytical scope to several citizen-science projects (cf. Brossard et al. 2005).

Behind these problems and contexts, the present contribution concluded on how the growing amount of citizen science projects and programmes could be accounted for their role in the scientific programming landscape, their role in science-society relations and the public understanding of science and with that how they can be treated by evaluators in the future. Potential implications for evaluation study planning, set-up, methodological choices and analysis of the results of traditional research performance indicators can now be discussed. Since the reflections are based on the experiences of two example evaluation cases in the Austrian context, opportunities and problems of scalability to a broader European or global context have to be taken into consideration.

REFERENCES

Birke, B., K. Fettelschoß and C. Gerloff-Gasser (2014): „Sparkling Science - Analyse von bildungsseitigen Auswirkungen", study for the Austrian Federal Ministry for Science, Research and Economy (BMWFW), http://www.fteval.at/de/evaluation_studies/all/2014/1724/

Birke, B. (2013): „Evaluierung des Förderprogramms Sparkling Science", study for the Austrian Federal Ministry for Science, Research and Economy (BMWFW), http://www.fteval.at/de/evaluation_studies/all/2013/1583/

Bornmann, L. (2012): "Measuring the societal impact of research", EMBO reports 13, Nr. 8 (10. Juli 2012): 673–76. doi:10.1038/embor.2012.99.

Brossard, D., B. Lewenstein and R. Bonney (2005): „Scientific knowledge and attitude change: The impact of a citizen science project". International Journal of Science Education 27, Nr. 9 (2005): 1099–1121. doi:10.1080/09500690500069483.

Kieslinger, B., T. Schäfer und C.M. Fabian (2015): "Kriterienkatalog zur Bewertung von Citizen Science Projekten und Projektanträgen", paper for the Austrian Federal Ministry for Science, Research and Economy (BMWFW), available at: https://www.zsi.at/object/publication/3864/attach/Kieslinger_Schaefer_Fabian_CS_Kriterien_2015.pdf

Laredo, P. (2007): "Revisiting the third mission of universities: Toward a renewed categorization of university activities?" Higher Education Policy, 20(4), 441-456.

Manahl, C., S. Dobner, T. Holocher-Ertl and B. Kieslinger (2015): „Sondierungs- und Evaluationsstudie Young Science - Zentrum für die Zusammenarbeit von Wissenschaft und Schule", study for the Austrian Federal Ministry for Science, Research and Economy (BMWFW), http://www.fteval.at/de/evaluation_studies/all/2015/1725

Manahl, C., I. Wagner and K. Schuch (2016): „Evaluierung des Wissenschaftlichen Impacts Von Sparkling Science", study for the Austrian Federal Ministry for Science, Research and Economy (BMWFW).

AUTHORS

ISABELLA WAGNER
Centre for Social Innovation
Linke Wienzeile 246, 1150 Vienna
Austria
wagner@zsi.at
+43 1 49 50 442 33

CAROLINE MANAHL
Centre for Social Innovation
Linke Wienzeile 246, 1150 Vienna
Austria
manahl@zsi.at
+43 1 49 50 442 18

SUSANNE DOBNER
Centre for Social Innovation
Linke Wienzeile 246, 1150 Vienna
Austria
dobner@zsi.at
+43 1 49 50 442 77

PROGRAMME EVALUATION AND ORGANISATIONAL DEVELOPMENT FOR TRANSDISCIPLINARY RESEARCH

GO YOSHIZAWA, NIKA ANDO AND KEIICHIRO TAHARA

INTRODUCTION

Transdisciplinary research attempts to integrate knowledge in a systemic way, transcend the scopeof disciplinary views and focus on problem solving of the life-world (Alvargonz‡lez 2011; Klein 2010). For the last decade, there have been theoretical and empirical discussions on transdisciplinary research evaluation (Wickson, Carew & Russell 2006; Walter et al. 2007; Garner et al. 2013; Belcher et al. 2016). In practice, a number of different transdisciplinary research programme management and evaluation activities globally come to the forefront, such as RuralEconomy and Land Use Programme (RELU) in the UK, Nesta in the UK, Mistra in Sweden, Superfund Research Programme in the US, and REPERE Programme in France. This bringspractical issues on how to discuss and assess the relative value of the range of research outputsthrough in-depth, multi-method analysis of knowledge integration and societal impacts (Koier &Horlings 2015; Ruegg & Thomas 2011).

Transdisciplinary research makes the boundary between academia and society in knowledgegeneration more blurred (Pohl 2008; Mobjšrk 2010). Not only knowledge producer and user, butalso other actors like sponsor, client and addressee perform key roles in evaluation of intermediary and its associated programmes (Yoshizawa & Nishimura 2013). Under the circumstances, knowledge exchange between actors appears more crucial for funding agencies to supporttransdisciplinary research as in the case of RELU (Phillipson, Lowe, Proctor & Ruto 2012) and Artand Humanities Research Council (AHRC) in the UK (Moreton 2016).

A dual challenge for public research funding agencies is meeting the demand for measuring and evaluating research performance while ensuring the wellbeing of research cultures and academic systems (S‡, Kretz & Sigurdson 2013). Given the gap between conventional metrics and the complexity of transdisciplinary research, evaluation becomes constructive (Klein 2008) andadopts the 'productive interactions' approach, which is contextual and oriented to process, learning and improving in relation to the daily activities of both researchers and stakeholders (Spaapen & van Drooge 2011). Where transdisciplinary research process is influenced by stakeholder participation (Angelstam et al. 2013), systematic involvement of a broad group of stakeholders, particularly policy makers and programme managers, from the very beginning in evaluation affectsthe utilization of evaluation results (Teirlinck et al. 2013). This also makes evaluation moreresponsive to wide range of stakeholders and synergistic with interactive policy approaches and theaim of policy learning (Abma 2004).

RISTEX: A FUNDING AGENCY FOR TRANSDISCIPLINARY RESEARCH

Among a great number of public or private funding agencies for research, technology and innovation in the world, only a few organisations dedicate transdisciplinary research and socialinnovation. One of them is the Research Institute of Science and Technology for Society (RISTEX), a public funding agency under the Japan Science and Technology Agency (JST). RISTEX aims to facilitate R&D for the creation of societal and public values with a variety of actors. Its focus is on challenge-driven innovation and societal embedding of R&D outcome with the concept of 'socio-technology', three key functions of which are (1) interaction and collaboration, (2) comprehensibility and interrelatedness, and (3) problem solving (Tahara, Yarime & Yoshizawa 2009).

In agenda setting and programme design, RISTEX first conducts horizon scanning in collaboration with a wide range of experts, stakeholders and citizens as potential programme users(Amanatidou et al. 2012). User involvement in the early stages of programme design can legitimately build capacity for research and innovation in strategic areas (Hessels et al. 2014). Thislargely reflects the spirit of open evaluation as an ongoing post-publication process of transparentpeer review and rating of papers (Kriegeskorte, Walther & Deca 2012). Once the programme area is specified, RISTEX then nominates a programme director, a programme officer and programmeadvisors, and calls for project proposals. The programme management team organises meetings, annual retreats, site visits, outreach activities and other formal or informal interactions with projectmembers. The past programme focus ranges from child safety, community resilience, intergenerational sustainability, aging society, service science, science of science policy, to science, technology and humanity.

PROGRAMME EVALUATION AT RISTEX

RISTEX's hands-on management makes it difficult to evaluate an R&D programme, not only because societal impacts of the programme can never be easily identified and assessed but also because it closely relates to RISTEX's own institutional issues as well as individual projects' own efforts. In 2013, outcomes of the R&D programme on science, technology and humanity wereanalysed and reported through interaction and collaboration between the programme managementteam, RISTEX officers and volunteered project members. The report delivered the programme'score findings and messages – that is, (1) linking between scientific knowledge and local knowledge,

(2) developing responsible experts, (3) drawing lessons from small societal challenges, and (4) building trust-responsiveness between actors (RISTEX 2013).

In the same year, RISTEX also formulated a new action plan based on an external reviewreport, in which three activities are proposed: (1) improving in-house analytical functions; (2) developing programme structures highlighting a story about problem solving; and (3) reforming theevaluation system. The last point is theoretically endorsed by a case study applying Beer's viablesystem model (VSM) that diagnoses the organisational structure of RISTEX (Tahara & Takahashi2014).

According to the above action plan, in 2015, RISTEX launched the Steering and Evaluation Committee (SEC) to conduct evaluation of R&D programmes and organisational management. Thecommittee consists of eight experts, including policy analysts, a private foundation programmeofficer, an ex-government officer and ex-project leaders funded by RISTEX. SEC developed a newformat for mid-term and ex-post programme evaluation. This format asks a programme manager to:

(1) identify problem subject to the programme and narrates a story about problem solving; (2) describe programme management and activities (process); (3) demonstrate progress to the goal(outcome); (4) specify output additionality (cf. HyvŠrinen & Rautiainen 2007), that is, what other programmes could not have performed (relevance); and (5) provide recommendations to RISTEX. This activity has an affinity for different approaches to story-based evaluation using logic models(McLaughlin 1999), qualitative case studies (Costantino & Greene 2003) and success stories (Dart& Davies 2003). Achieving consistency within and between individual programmes that bring ashared vision through the story, it may improve organisational evaluation capacity (Cousins, Goh, Elliott & Bourgeois 2014) and organisational development (McClintock 2004).

SEC also tried to reform the evaluation system and make mid-term evaluation more relevantand effective. In the past, mid-term evaluation was conducted only to perform public accountability but not to affect programme governance (including project selection). The committee launched an informal meeting with each programme governance board (programme manager and administrators, external programme advisors) at the early stage, that is, 6 months to 1 year after the start of the programme, aiming for information exchange and consultation for programme development. Evaluation thus becomes more formative, interactive and constructive.

TYPOLOGY OF KNOWLEDGE AND ACTORS

The above kind of programme evaluation activities reflexively problematizes knowledge and actorsfor transdisciplinary research. Inspired by early observations (Pohl 2008; Mobjšrk 2010;Yoshikawa 2013), this study identifies three types of knowledge for transdisciplinary research bycaricaturing researchers and practitioners of each type (Table 1).

Type of knowledge	Researchers	Practitioners
Observational	Observe society and nature to providegeneralized knowledge; have interest and responsibility in establishing and maintaining academic discipline.	Observe society and nature to grasp regional needs and social problems; advocate for solving the problemsidentified from local knowledge.
Synthetic	Formulate a methodological and institutional model for problem solving; have interest and responsibility in co-production of knowledge with wider stakeholders.	(Personally) hold design thinking and network to depict problem environmentand solution paths; have interest in problem solving and manage stakeholders.
Socially contributive	Offer knowledge originated at themselves and adjusted through interaction with society; commit to society and nature for their social responsibility	Provide a solution to the given problem by action based on tacit knowledgecoming from their own experience and idea.

Table 1. Type of knowledge and actors for transdisciplinary research

Past studies on transdisciplinarity and its related concepts are likely to downplay the role ofsynthetic knowledge whilst emphasising a bridge between observational knowledge and socially contributive knowledge. Synthetic knowledge is situated, reflexive and anticipatory and illuminatescomprehensibility and interrelatedness of sociotechnical systems and transformations.

ORGANISATIONAL DEVELOPMENT

Highlighting actors who demonstrate synthetic knowledge cannot be limited at the programme levelbut should also be addressed at the organisational level, as necessary in participatory, flexible and responsive evaluation (Boyd et al. 2007). In the case of RISTEX, one of the organisational challenges is oriented to responsible reform of the research and innovation ecosystem. The reason why synthetic researchers remain relatively few despite the increasing societal needs of transdisciplinary research is deeply originated at the culture of research communities. For this,academia is a society in which problems are to be solved through transdisciplinary activities. Another challenge for RISTEX is in policy arena. Public policymaking and implementation isexpected as a useful solution for many of social problems. However, it is a major hurdle for individual projects (and even programmes) to approach policymakers and support policymaking. Asa governmental agency, RISTEX now faces how to develop effective formal and informal channelsfor the national and regional policy process by integrating project or programme outcome.

Transdisciplinary research requires a long-term commitment over a

5-to 10-year time framewhile accommodating different expectations and values of participants and fostering social learning (Roux et al. 2010). This necessarily raises issues on participatory management of funding agency for transdisciplinary research. Stakeholder and public engagement in organisational managementcan be found at the programme development, the project design and the project levels. Publicengagement in project selection, as observed in the UK (Rowe et al. 2010), would be one of further inclusive strategies for RISTEX. In order to conduct and utilise evaluation more effectively, RISTEX may have to consider stakeholder participation in the communication of results phase ofprogramme evaluation (Greene 1988), in terms of the internationalisation of funding agencies(Reale et al. 2012).

CONCLUSION

Recent efforts in the reform of programme evaluation at RISTEX do not merely reveal the difficulty of evaluation on transdisciplinary research and social innovation but also require the needs oforganisational development by broadening out evaluation with wider participants (cf. Ely, Van Zwanenberg & Stirling 2014) and reflexively arranging knowledge and actors.

REFERENCES

Abma, T. (2004): "Responsive evaluation: the meaning and special contribution to public administration," Public Administration, vol. 82, pp. 993-1012.

Amanatidou, E. et al. (2012): "On concepts and methods in horizon scanning: lessons frominitiating policy dialogues on emerging issues," Science and Public Policy, vol. 39, pp. 208-221.

Angelstam, P. et al. (2013): "Solving problems in social-ecological systems: definition, practice and barriers of transdisciplinary research," AMBIO, vol. 42, pp. 254-265.

Alvargonz‡lez, D. (2011): "Multidisciplinarity, interdisciplinarity, transdisciplinarity, and the sciences," International Studies in the Philosophy of Science, vol. 25, pp. 387-403.

Belcher, B. M., K. E. Rasmussen, M. R. Kemshaw and D. A. Zornes (2016): "Defining and assessing research quality in a transdisciplinary context," Research Evaluation, vol. 25, pp. 1-17.

Boyd, A. et al. (2007): "Systemic evaluation: a participative, multi-method approach," Journal of the Operational Research Society, vol. 58, pp. 1306-1320.

Costantino, R. D. and J. C. Greene (2003): "Reflections on the use of narrative in evaluation," American Journal of Evaluation, vol. 24, pp. 35–49.

Cousins, J. B., S. C. Goh, C. J. Elliott and I. Bourgeois (2014): "Framing the capacity to do and useevaluation," in J. B. Cousins and I. Bourgeois (Eds.): Organizational capacity to do and use evaluation. New Directions for Evaluation, vol. 141, pp. 7-23.

Dart, J. and R. Davies (2003): "A dialogical, story-based evaluation tool: The most significant change technique," American Journal of Evaluation, vol. 24, pp. 137–155.

Ely, A., P. Van Zwanenberg and A. Stirling (2014): "Broadening out and opening up technology assessment: approaches to enhance international development, co-ordination and democratisation," Research Policy, vol. 43, pp. 505-518.

Garner, J., A. L. Porter, M. Borrego, E. Tran and R. Teutonico (2013): "Facilitating social and natural science cross-disciplinarity: assessing the human and social dynamics program,"Research Evaluation, vol. 22, pp. 134-144.

Greene, J. C. (1988): "Communication of results and utilization in participatory program evaluation," Evaluation and Program Planning, vol. 11, pp. 341-351.

Hessels, L. K., T. Wardenaar, W. P. C. Boon and M. Ploeg (2014): "The role of knowledge users in public-private research programs: an evaluation challenge," Research Evaluation, vol. 23, pp. 103-116.

HyvŠrinen, K. and A-M. Rautiainen (2007): "Measuring additionality and systemic impacts ofpublic research and development funding -the case of TEKES, Finland," Research Evaluation,vol. 16, pp. 205-215.

Klein, J. T. (2008): "Evaluation of interdisciplinary and transdisciplinary research: a literature review," American Journal of Preventive Medicine, vol. 35, pp. S116-S123.

Klein, J. T. (2010): "A taxonomy of interdisciplinarity," in R. Frodeman, J. T. Klein and C. Mitcham (Eds.): The Oxford Handbook of Interdisciplinarity. Oxford University Press.

Koier, E. and E. Horlings (2015): "How accurately does output reflect the nature and design oftransdisciplinary research programmes?" Research Evaluation, vol. 24, pp. 37-50.

Kriegeskorte, N., A. Walther and D. Deca (2012): "An emerging consensus for open evaluation: 18 visions for the future of scientific publishing," Frontiers in Computational Neuroscience, vol. 6, Article 94.

McClintock, C. (2004): "Using narrative methods to link program evaluation and organization development," Evaluation Exchange, vol. 9, no. 4, pp. 14-15.

McLaughlin, J. A. (1999): "Logic models: a tool for telling your program's performance story,"Evaluation and Program Planning, vol. 22, pp. 65-72.

Mobjšrk, M. (2010): "Consulting versus participatory transdisciplinarity: a refined classification oftransdisciplinary research," Futures, vol. 42, pp. 866-873.

Moreton, S. (2016): "Rethinking 'knowledge exchange': new approaches to collaborative work in the arts and humanities," International Journal of Cultural Policy, vol. 22, pp. 100-115.

Phillipson, J., P. Lowe, A. Proctor and E. Ruto (2012): "Stakeholder engagement and knowledgeexchange in environmental research," Journal of Environmental Management, vol. 95, pp. 56-65.

Pohl, C. (2008): “From science to policy through transdisciplinary research,” Environmental Science & Policy, vol. 11, pp. 46-53.

Reale, E., A. Inzelt, B. Lepori and P. van den Besselaar (2012): “The social construction of indicators for evaluation: internationalization of funding agencies,” Research Evaluation, vol. 21, pp. 245-256.

Research Institute of Science and Technology for Society (RISTEX) (2013): Expanding the Rangeof Participants and a New Role for Experts: Interaction between Science, Technology and Society. Report of R&D Focus Area: Science, Technology and Humanity. Available athttps://www.ristex.jp/result/science/pdf/20131128_2.pdf

Roux, D. J., R. J. Stirzaker, C. M. Breen, E. C. Lefroy and H. P. Cresswell (2010): “Framework for participative reflection on the accomplishment of transdisciplinary research programs,” Environmental Science & Policy, vol. 13, pp. 733-741.

Rowe, G., D. Rawsthorne, T. Scarpello and J. R. Dainty (2010): “Public engagement in research funding: a study of public capabilities and engagement methodology,” Public Understanding of Science, vol. 19, pp. 225-239.

Ruegg, R. and P. Thomas (2011): “Tracing from applied research programs to downstreaminnovations: value in multiple techniques,” Research Evaluation, vol. 20, pp. 305-312.

S‡, C. M., A. Kretz and K. Sigurdson (2013): “Accountability, performance assessment, and evaluation: policy pressures and responses from research councils,” Research Evaluation, vol. 22, pp. 105-117.

Spaapen, J. and L. van Drooge (2011): “Introducing ‘productive interactions’ in social impact assessment,” Research Evaluation, vol. 20, pp. 211-218.

Tahara, K. and S. Takahashi (2014): “Evaluation system to ensure the viability of funding agencies:development and validation of viable system model for intermediary,” unpublished.

Tahara, K., M. Yarime and G. Yoshizawa (2009): “The concept of ‘sociotechnology’ and funding agencies dedicated to science and technology for society,” presented at Atlanta Conference onScience and Innovation Policy.

Teirlinck, P., H. Delanghe, P. Padilla and A. Verbeek (2013): “Closing the policy cycle: increasing the utilization of evaluation findings in research, technological development and innovation policy design,” Science and Public Policy, vol. 40, pp. 366-377.

Walter, A. I., S. Helgenberger, A. Wiek and R. Scholz (2007): “Measuring societal effects of transdisciplinary research projects: design and application of an evaluation method,” Evaluation and Program Planning, vol. 30, pp. 325-338.

Wickson, F., A. L. Carew and A. W. Russell (2006): “Transdisciplinary research: characteristics, quandaries and quality,” Futures, vol. 38, pp. 1046-1059.

Yoshikawa, H. (2013): “Service innovation in the age of sustainability,” in Y. Shimomura and K. Kimita (Eds.): The Philosopher’s Stone for Sustainability: Proceedings of the 4th CIRPInternational Conference on Industrial Product-Service Systems, Tokyo Japan, November8th-9th, 2012. Springer Berlin Heidelberg.

Yoshizawa, G. and Y. Nishimura (2013): “The role of intermediaries linking knowledge and socialand public values,” IAMOT 2013 Proceedings, #1569691229.

AUTHORS

GO YOSHIZAWA
Corresponding author
Graduate School of Medicine, Osaka University
2-2 Yamadaoka, Suita, Osaka 565-0871
Japan
Tel: +81-6-6879-3688
E-mail: go@eth.med.osaka-u.ac.jp

NIKA ANDO
Research Institute of Science and Technology for Society (RISTEX)
5-3 Yobancho, Chiyoda-ku, Tokyo 102-8666
Japan

KEIICHIRO TAHARA
Institute for Future Engineering
2-6-11 Fukagawa, Koto-ku, Tokyo 135-8473
Japan

ASSESSING THE IMPACT OF RESEARCH IN SOCIAL SCIENCE AND HUMANITIES: A COMPARATIVE PERSPECTIVE ON NATIONAL EVALUATION SYSTEMS IN FRANCE, SPAIN AND GERMANY

LUCIO MORETTINI, EMANUELA REALE, TERESA SORDÉ AND ESTHER OLIVER

BACKGROUND AND AIM OF THE PAPER

Assessing the social impact of research is becoming more important in the debate on evaluation policies. The diffusion of the neo-liberal paradigm (NPM) and the changes in the concept of research and its results led toward the emergence of efficiency, effectiveness, and productivity driving the way in which the activities must be managed. Accountability and the principle of value for money further enforce the mentioned trend, as well as the establishments of standards as benchmarks of successful performance (Brunsson and Jackobsson, 2002).

This process of change goes with the cutting of public resources (first and foremost funding) devoted to R&D, and with the emergence of a quest from the policy makers on behalf of society about the utility of research and its capability to contribute to the progress and well-being of the whole community. The needs of 'evidence-based justification' to sustain R&D through public funding, and the push toward focusing on "relevant" themes of investigation affected also the reflexivity about the public investment on R&D, questioning about the 'excellence' of the research produced, the capability of the research to address grand challenges for the sustainable development and innovation, to generate breakthrough and innovation, and definitely the impact produced by the R&D activities on science, society, economy and policy (Penfield et al., 2014).

Despite the interest, impact evaluation of R&D is strongly affected by the time lag and attribution problems, and solutions elaborated to solve them are still striving debates among scholars and policy makers about their capability to contribute to the evidence-based policy process (Hughes, A. and B. Martin, 2012; Spaapen et al., 2014; Reale et al. 2014).

The paper investigates how the evaluation of impact is implemented in national R&D systems, trying to understand how the configuration of the research systems in different countries can influence perception and ideas of social impact of research in Social Science and Humanities among insiders (researchers, policymakers and research managers). The analysis is devoted to control whether the research evaluation systems: i) use (or intend to use) mechanisms to evaluate the social impact from the research results, ii) gather (or try to gather) evidence about the social impact of the projects assessed and make it visible, iii) include indicators for the impact assessment.

The paper takes a comparative perspective of three case studies about Germany, Spain and France. All of them are large countries of continental Europe that have suffered less than other instances of the reforming of New Public Management but they present also deep differences about the role and autonomy of Higher Education Institutions and the structure and the role in research policy of central and local administration.

The focus on SSH allows to shed light on fields whose contribution to economy and society is questioned, and considered less relevant than natural sciences; SSH are sometimes perceived as a sort of non-productive investment, whose added value is difficult to identify. Some authors claimed about Arts and Humanities perceived as "useless frills" (Nussbaum, 2010) and often unknown ("forgotten sciences", Bod, 2013). Moreover, the possibility to measures the outputs of the research activities in these fields of science is constrained by the lack of adequate indicators and metrics (Ochsner et al., 2016). Problems of time lag and attribution are particularly high when the assessment of SSH results is concerned, as well as the identification of the relevant stakeholders.

CONCEPTUAL AND THEORETICAL FRAMEWORK

The political system and the administrative culture and traditions play a central role structuring the research and HE systems of the countries: for instance Bleiklie and Michelsen (2013) found that influence of political system can be more ambiguous and flexible but also deeper than expected, while Whitley (2007) shows how organized groups of politician, business elites and other policy advisers could strongly condition research activity if there is not separation with national academic system. The analysis of research systems' configuration in European countries also shows a large heterogeneity among national R&D systems that are characterized mainly for peculiarities of intermediary organizations between policymakers and researchers. The characteristics of intermediaries are not neutral with respect to the structures of research system and to its history as reported by Ferlie et al. (2009).

Speaking about political systems, two basic dimensions of the state structure are considered: the first is the vertical dispersion of authority, which is the case of decentralized federal countries where the power is delegated to sub-national entities, thus being less uniform than centralized countries where the power is owned at central level and the focus on delivery and results is more pronounced (Bleiklie and Michelsen, 2013). The second dimension is the horizontal coordination at central government level. As to the national cases in this paper, Pollitt & Bouckaert (2003) consider France a country more coordinated than Germany because the administrative élites of officers balance the fragmentation of the system, while in Germany the fragmentation is higher because the mechanism of coordination is supposed to be less efficient. In this context, Spain represents a peculiar case: a country with a strong tradition of central administration that during the last decades has found an equilibrium with the demands of larger administrative autonomy of local communities.

When the paper refers to administrative culture and traditions, it indicates two typologies used by Peters (2008) and Painters and Peters (2010). One is the Napoleonic tradition, where the focus is on law -as a mean of the state to intervene on society, and on administration -strictly related to the laws, with a small role that in principle the societal actors are supposed to play. However, since the presence of implementation gaps is a key characteristic of this tradition, distances between what is prescribed by law and the actual existence or utilization of management tools are always observed. The other typology is the German tradition, with a dominant role of the state to integrate the different parts of the system, and a strong role of the bureaucracies to assure the compliance to regulations and rules.

The paper assumes that since the political – administrative structure influences the configuration of HE system in each country (Bleiklie and Michelsen, 2013) it is likely to affect the research system as well. More specifically, the configuration of the state structure (centralized-decentralized structure, coordinated-fragmented) and the administrative culture and traditions of the countries (Napoleonic vs German) shape the implementation of the evaluation of R&D impact as to:

a. The importance given to the ex-ante R&D impact assessment vs the ex-post one. The latter is more related to the administration implementing type of performance-based approach, with a relevance of the central control over the activities and results of the performer.
b. The actors involved in impact assessment, and the autonomy they have with respect to the state and the performers. Here we expect different roles of the intermediaries to implement the impact assessment and different spaces of maneuver of the performers with respect to the state steering;
c. The indicators and metrics used, with a higher prominence of metrics and indicators in countries with a centralized political system a law-driven administrative culture;
d. The association between the R&D impact assessment and the R&D funding. This item is related to the specification of the concept of impact in different system, and to the use of evaluation in HE policy.

The case studies analyzed in this paper present the following characterization.

Germany is a federal state where regional governments play a central role in S&T policies. Government intermediary level is characterized by important actors such as DFG, which manage the most important funding instrument. The performer level presents a strong division of the roles: research system has a double and parallel binary between universities and public research organization. Both of them preserve their autonomy with rigid self – assessment, however their political counterpart is different: universities have the supervision of local government of federal states (Lander), while PRO are supervised directly by central government that has also a soft power of coordination of research policy of each Land. The most significant implementation of New Public Management principles is represented by a soft budget constraint that allows evaluation to have a limited influence.

Administrative and political system in France has a Napoleonic imprint and research system, even preserving its autonomy respect to policymakers, is part of this administrative structure, thus it is characterized by a strong centralization. The French research system has suffered less than others the influence of NPM, however the reforming instances did not remain unimplemented. Given the emergence of new social actors and the will of local institutions to carve out a decision space, French research system has been involved in in a network governance narrative. Respects to Germany, there are more opportunities to exchange between HEIs and social and political actors until the possibility to have a common vision in the definition on research policy. Two agencies act as intermediaries for R&D funding instruments and evaluation.

Spain too presents a Napoleonic structure, but, as said before, during the last years some reforms occurred to the administrative and political system, with the attribution of larger autonomies to local communities. This process fuelled the establishment of new actors in the administrative area that have the possibility to contribute to national policy in several fields, including research. About research, the aim was to incentive the integration of universities and research centers in local socio-economic systems, with a supervision of central government that could guarantee homogeneity and unity. However, the result is quite different: several local authorities have created a regional research system that duplicates the national one. Currently, Spanish research system is the sum of national and local research systems that act in parallel with few points of contact but each one with its own funding, evaluation authority and research policy objectives.

Given these differences, we expect that in Germany decision makers and academics could be more inclined to see a strong distance between assessing the effects of research work inside the research system –either the scholar community or the organizations, and outside the research system. While scholars can pursue and assess the former, the latter are not in the disposal of researchers, thus political and social impact is not an issue at stake. On the other hand, French decision makers and academics could be more likely to consider research as something that is fully integrated in society, with a "natural" transmission of the results to other members of the political and administrative system with which they share the determination of the path of research policies. Implementation nonetheless very often shows a very limited importance of research impact assessment. Finally Spanish research system on one hand seems to facilitate possibility to comparisons between researchers, research managers and policymakers because of the multiplication of actors involved in research policy. On the other hand, the fragmentation of research in several parallel research systems seems to guarantee the exchange only between actors within the same system.

In all the countries the impact assessment of SSH should not show differences in comparison with other fields, but the identification of metrics adapt to capture the effects produced on society, with productive interactions and dialogues with stakeholders having higher importance than input/output approaches based on metrics.

Finally, the difficult to find a common definition for social impact of evaluation between elements of the same research system could be amplified by the influence that different configurations of research and political – administrative systems have on several topics of research policy.

EMPIRICAL FRAMEWORK AND DESCRIPTION OF THE METHODOLOGY

Testing the perceptions and ideas of actors with different role in the R&D system is the base for the comparative analysis of France, Spain and Germany. To do that we integrate the results coming from desk analysis and interviews to eight representatives per country[1], selecting them among following categories:

1. Research Evaluator/Peer Reviewer for the purposes of funding and/or accountability in charge of SSH evaluation;
2. Research Manager/Scientific Officer within a research funding agency (within a government department or in an exclusive funding body quango);
3. Policymaker/politician/budget holding civil servant;
4. Academics.

The interviews are focused on the current work of representatives, policy makers and reviewers or researchers about the evaluation of research results, with a deep investigation on tools and technics. Furthermore, was asked to the subjects involved to define the social impact of research, providing examples of the effects observed on people's lives and on policy actions.

The goal of these interviews is to collect qualitative that could show us the vision of different actors about the same topic, trying to identify common characteristics and strong differences between countries and among them.

Desk analysis deals with the following:

√ Legal and policy documents approved by national parliaments on SSH research evaluation or national research policy;

√ Documentation related to recent calls for research projects, and other relevant documents of the process of selection.

PRELIMINARY RESULTS

Preliminary results of our analysis show an opposite approach to the social impact of research between Germany, Spain and France.

In Germany, social impact is seen as a natural part of research in SSH: results will have an effect on society and legislation even if in different way and with different timing. In this context, academics are surer about the role of social impact than policymakers, which appear more prudent. On the other hand, in France insiders tend to demarcate difference between divulgation of research results and social impact. According to this vision, the second one has a negligible role and if policy presents loanable items from research results, these elements are not a clear social impact effect but most likely the result of "coincidences". This point of view seems to be common to the French insiders even if, especially among policymakers, seems to grow the interest on a more complete analysis of this aspect of research. About Spain subjects involved in interviews gave a more elusive definition of social impact; however, regularity among the answers about the role of its evaluation can be outlined. University representatives see in the social impact evaluation a useful element to define the purpose of research in SSH. On the other hand, representatives of local research assessment institutions are less interested in social impact evaluation, defining it as something unrelated to research evaluation and that cannot influence funding allocation. Finally evaluators from national institutions seems to have a tougher stance toward the social impact assessment, defining social impacts something of external to research, so unnecessary to evaluate with the scientific impact of research.

The results achieved underline how the debate on impact is focused not only on its quantification but also on its specification. More and more policymakers are interested to understand a) the effect of research on the rest of research policy or on the productive system, and b) if there is a transfer of knowledge to the society. However, a distance between the ideas of social impact of policymaker and the researchers' point of view on the topic might appear. This distance is more pronounced in France than in Germany, a fact confirming the capability of R&D systems in decentralized and fragmented political countries to avoid the problems of implementation gap, which is indeed a feature emerging in the centralized and unitary country of France as far as the R&D impact is concerned. In this context, Spain is an interesting and peculiar case: fragmentation of research system seems lead to situation where French result is reproduced for each one of the parallel research systems present at regional level.

REFERENCES

Bleiklie I. and S. Michelsen (2013): "Comparing HE policies in Europe: structures and reform output in eight countries" Higher Education 65 (2013), pp. 113 – 133

Bod R. (2013): A New History of the Humanities. Oxford University Press UK

Brunsson N. and Jackobsson B. (2002): A World of Standards. Oxford University Press, UK

1 Impact – EV is an EUFP7 Project coordinated by the University of Barcelona based on the comparison of thirteen national research systems, nine from Europe (as well as Germany and France are involved United Kingdom, Netherlands, Ireland, Hungary, Sweden and Spain) and four from the rest of the world (USA, Brazil, Hong Kong and Australia).

de Jong et al. (2011): "Evaluation of Research in Context: an approach and two cases" Research Evaluation, 20 (1), 61-72

Ferlie E., C. Musselin, G. Andresani (2009): "The governance of hgher education systems: a public management perspective" in C. Paradeise, E. Reale, I. Bleiklie, E. Ferlie. University Governance – Western European comparative perspectives. Springer, pp. 1 – 19

Hughes, A. and B. Martin, (2012): 'Council for Industry and Higher Education, UK Innovation Research Centre. Enhancing Impact. The Value of Public Sector R&D', Summary report, (Pubd online) accessed 26 Oct 2012.

IMPACT-EV (2015): WP2 Final Report. IMPACT-EV

Nussbaum M. C. (2010): Not For Profit. Why Democracy Needs the Humanities. Princeton University Press

Ochsner M., Hug S. E., Daniel H-D (2016): Research Assessment in the Humanities. Springer Open, Switzerland

Painter M. and Peters B. (2010): Tradition and Public Administration. Palgrave Macmillan UK

Penfield T., M. J. Baker, R. Scoble and M. C. Wykes (2014): "Assessment, evaluations, and definitions of research impact: A review" Research Evaluation, 23 (2014) pp. 21–32

Peters B. G. (2008): The Politics of Bureaucracy. Routledge, London

Reale E., M. Nedeva M., D. Thomas and E. Primeri (2014): "Evaluation through impact: a different viewpoint" Fteval JOURNAL for Research and Technology Policy Evaluation, 39, pp. 36 – 41

Spaapen, J. & van Drooge, L. (2011): "Introducing "productive interactions" in social impact assessment" Research Evaluation, 20 (3), 211-218

Whitley R. (2007): "Changing governance of the Public Sciences: the consequences of establishing research h evaluation system for knowledge production in different countries and scientific fields" in R. Whtiley and J. Glaser. The changing governance of the sciences. Springer, Dordrecht pp. 3 – 27

AUTHORS

LUCIO MORETTINI
CNR- IRCRES Research institute on sustainable economic growth

EMANUELA REALE
Corresponding author
CNR- IRCRES Research institute on sustainable economic growth
emanuela.reale@ircres.cnr.it

TERESA SORDÉ
CREA – UB University of Barcelona

ESTHER OLIVER
CREA – UB University of Barcelona

SSH & THE CITY. A PROPOSAL TO MAP SOCIAL ENGAGEMENT THROUGH SOCIAL MEDIA AND WEB-LINK ANALYSIS[1]

NICOLAS ROBINSON-GARCIA, THED N. VAN LEEUWEN AND ISMAEL RAFOLS

INTRODUCTION

Current evaluation frameworks in research policy were designed to address: 1) life and natural sciences, 2) global research communities, and; 3) scientific impact. They do not adapt well to SSH scholarship, to local interests, or consider broader societal impacts. Their focus on outputs, implicitly assumes a linear relation between the activity (research) and the expected result (publication). This is particularly problematic to the SSH areas as they can address different audiences (Nederhof 2006). But many of the 'impacts' SSH activity may have in society are due to multiple factors (problem of attribution) and are of a secondary nature that is, the outcome may not be traceable to any single given output but to cascading effects (Upton et al. 2014). We propose a network approach for identifying societal contributions in local contexts. The goal of this approach is not to develop indicators for benchmarking, but to map interactions for strategic assessment.

We discuss three evaluation frameworks and propose a methodology to capture societal interactions between SSH researchers and their cities. Many countries are putting an increasing emphasis on the societal impact discourse when assessing research performance. They are calling for evidences of societal impact, urging researchers to engage on social outreach and public engagement.

CONCEPTUAL FRAMEWORKS

To avoid linear "impact" indicators, we suggest the application of alternative evaluation frameworks which consider process-based indicators. Because output-based approaches consider the relation between the activity and the outcome to be linear, it ties researchers to take an expected course of action, intruding in many cases in their communication patterns (Fuchs 2014).

Here we consider three frameworks which, to our understanding, define and describe in a consistent manner process-based interactions, but have so far failed at proposing a scalable and quantitative methodology to apply them. These frameworks intend to overcome problems raised when trying to assess scholarly performance and activities beyond scientific impact. The first approach we refer to is the 'third stream metrics' (Molas-Gallart et al. 2002) where they differentiate between activities and capabilities of universities. Capabilities are defined as of two types: physical facilities and knowledge capabilities. Based on this key differentiation a set of indicators is developed associated to each aspect. The second framework is based on the 'productive interactions' concept (Spaapen & van Drooge 2011), which adopts a process-based perspective. Its originality is due to explicitly driving away from an 'impact' discourse (understood as the effect of research in society).

These two frameworks could be analysed in terms of networks, making the following analogy. Capabilities represent nodes; and activities, defined as productive interactions, are the edges of a network by which academics and institutions interact and exchange knowledge with non-academics. Here is where the third framework, the 'knowledge value' framework (Rogers & Bozeman 2001), becomes relevant. This framework aims at defining the unit of analysis that should be used for evaluation instead of R&D programs. They present two core concepts: knowledge value collectives (KVC) and knowledge value alliances (KVA). In short, KVC is defined as the set of individuals who share a common knowledge base, while KVA is defined as a subset of a KVC where individuals are interacting with each other.

EXPECTED METHODOLOGICAL CONTRIBUTION

In our understanding, the three proposals referred to above describe the research process and collectives as interactions, flows and connections, but fail to recognize the network approach as a way of implementing them empirically. Instead, they rely on unidimensional indicators. By using social media and web-link analysis we can identify interactions between academics and local stakeholders. We consider that the power of these tools is not so much on understanding their meaning as 'acts' to develop impact or visibility metrics whenever a mention to a research article is made (Haustein et al. 2015), but as proxies for personal interactions. This methodology may benefit SSH areas, due to the direct and informal nature of relations between scholars from these fields and non-academics (Olmos-Peñuela et al. 2014). The strength of such approach is based on its potential to monitor unstructured interactions, as well as those which are the result of a specific research action (i.e., an R&D project). Therefore, research policy makers can monitor and better com-

1 Nicolas Robinson-Garcia is supported by a Juan de la Cierva-Formación Fellowship granted by the Spanish Ministry of Economy and Competitiveness.

prehend the process of interactions developed, and also identify other hot spots where productive interactions may well be taking place.

PRELIMINARY RESULTS AND FURTHER STEPS

In order to test our methodology, we will conduct several case studies in different cities of Spain and The Netherlands focusing for each of them on specific events and SSH areas. Here, we show two examples of the types of networks we aim at discovering through our approach. These examples are very preliminary results and should be considered as illustrations of the methodology rather than complete analyses of our case studies.

Figure 1 shows some preliminary results of relations between a sample of spin-offs, music-related and movie-related institutions in the city of Granada. As observed, although interactions through the web are not fully explored in this first approach, we can already capture ties between cultural events and institutions and the university. We also observe the strength of the tie between the city and the university, highlighting its role as an anchor institution for cultural life of the city (Goddard et al. 2014).

In our second test (figure 2) we focus in the case of the city of Valencia and web-links between university and a sample of local associations and cultural events. A complete different picture emerges here. First, we find three components instead of one: one formed by a neighbourhood association and a local association, a second one where the University of Valencia (uv.es) acts as central node connecting a local political association and other associations defending local agriculture, and a third component relating theatre and arts institutions and associations. Second, we observe how three of the four universities analysed (ucv.es, uchceu.es and upv.es) remain as isolated nodes of the network. In this case, the University of Valencia seems to be the only one establishing ties with local bodies, although its role in our network does not seem to be as central as in the former case. Interestingly, local authorities such as the local council (valencia.es) the regional department of culture (culturarts-generalitat.es) do not play a crucial role either.

Although these examples need further refinement, they offer a good example of the type of interactions we are proposing to capture through our network approach. In this case we have used an out-link analysis between a selected sample of institutional websites without going into much depth in our analyses. Future directions will go into the designing and analysis on the relation between the university and specific events or institutions in the city. Our aim is to go beyond an institutional perspective and make use of Twitter, blogs and Facebook to identify direct personal interactions that may reflect a greater (but also informal) role of universities in the case of Valencia for instance, than previously noted through our web-link analysis.

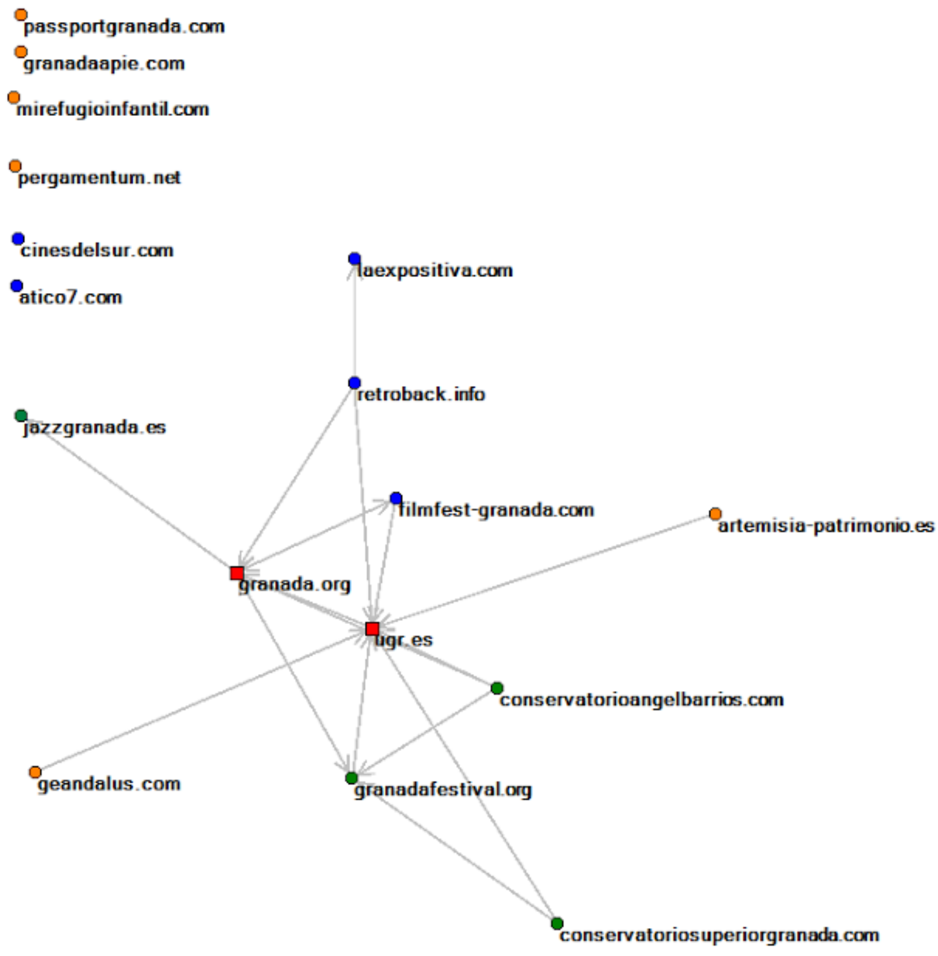


Figure 1 Out-link network of the University of Granada and its interactions with other local institutions. Node colours: red > University of Granada and City Council; green > music related festivals and institutions; orange > spin-offs; blue> movie related festivals and institutions. Depth of crawling: 1.

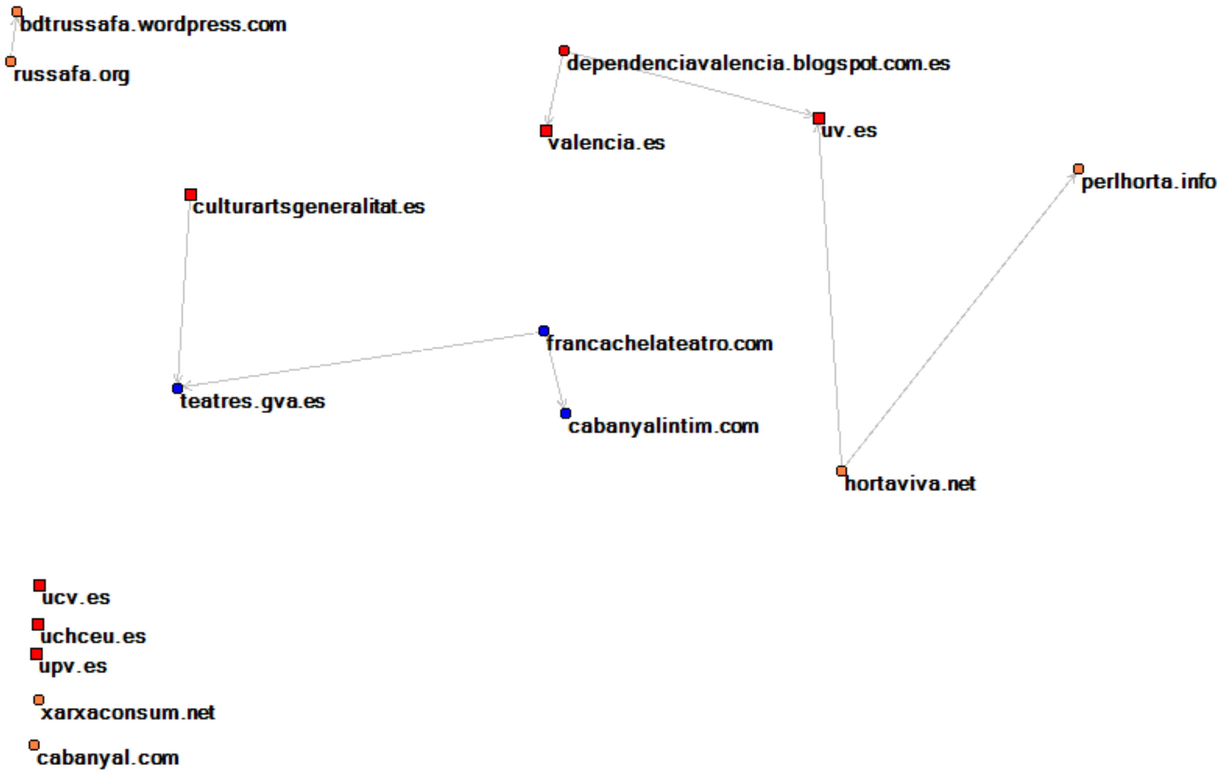


Figure 2 Out-link network of four Valencian universities and their interactions with other local institutions. Node colours: red > universities and local and regional public institutions; green > music related festivals and institutions; orange > social and neighbourhood associations; blue> theatre and arts related associations. Depth of crawling: 1.

In the case of Twitter data, we expect to encounter a series of challenges derived from the nature of this social media platform. These are the following:

1. Twitter data is retrieved by querying its API, which means that our network will be dependent on such query and we will not know to what extent we are showing a complete picture of the activity we are querying.
2. Because of such difficulty to retrieve and manage complete datasets, we cannot analyse local interactions between academics and non-academics in general, but must focus on specific events or users.
3. Due to the size of the network and the informal nature of the platform, nodes representing academics are not self-evident or easily identifiable.

However, Twitter has the positive aspect of informing us of different types of ties between nodes, hence links can be established in terms of followers and followees, mentions or retweets. Such distinctions allow us to distinguish between social distance and network paths (Watts, Dodds & Newman, 2002). That is, two dimensions of the network which could allow us to identify 1) potential academics, and 2) potential actors involved in a given event. We define social distance as that related to the acknowledgement two nodes make of each other. Hence, if node A and node B follow each other, we consider them 'socially' close to each other. Network paths are defined as those which link two nodes by discussing common topics.

Academic nodes are identified in terms of social distance. This can be done following two possible strategies:

1. We query for a given university's Twitter account and identify all its followers and followees. We consider academics all those nodes which follow the university and are followed by the university.
2. We identify through Altmetric.com API all Twitter accounts discussing research papers authored by a given university. We then cross these accounts with the university account and consider those nodes to be local and academic. This definition would be more restrictive.

CONCLUDING REMARKS

The present paper proposes a novel approach for analysing societal contributions of SSH academics to their cities. Although the methodology is not fully implemented, very preliminary results are offered for illustrating its potentialities. This proposal shows a quantitative approach which can be of use for research policy makers. Such methodology is characterized by the following aspects:

1. **Local versus global impact.** It is directed at the contribution academics make to local development, as opposed to traditional indicators such as citation metrics.
2. **Sociocultural impact of academia.** Societal relevance is traditionally considered in socioeconomic terms. Focusing on interactions rather than impact-based indicators offers a wider

perspective as to what is considered 'societal'.

3. **Social networks instead of impact indicators.** As illustrated by the three evaluation frameworks presented above, recent developments in research evaluation are directed at analysing interactions between institutions, academics and other actors (firms, non-academics, society in general). But they fail at operationalising their proposals in terms of network analysis. Network analysis and mapping have already been proven a powerful tool for research evaluation (Noyons, 2005), however they have not yet been applied in the context of societal impact.
4. **Social media as a data source.** Studies analysing societal contributions in SSH usually recollect their data either through surveys (Olmos-Peñuela, Castro-Martínez & D'Este, 2014) or interviews (Olmos-Peñuela et al., 2014). Such methods are of limited success for the following reasons: 1) they are costly in terms of time and money, 2) they are highly dependent on the subject's capacity to inform fully of their activity, and 3) they are intrusive, intervening with academics' activities.

REFERENCES

Fuchs, M.Ž. (2014). Bibliometrics: Use and abuse in the humanities. Blockmans W., Engwall, L., & Weaire D. (eds) Bibliometrics. Use and Abuse in the Review of Research Performance, pp. 107–16. Portland Press: London.

Goddard, J., Coombes, M., Kempton, L., & Vallance, P. (2014). Universities as anchor institutions in cities in a turbulent funding environment: vulnerable institutions and vulnerable places in England. Cambridge Journal of Regions, Economy and Society, 7(2), 307–25.

Haustein, S., Bowman, T.D., & Costas, R. (2015). Interpreting 'altmetrics': viewing acts on social media through the lens of citation and social theories. arXiv:1502.05701.

Molas-Gallart, J., Salter, A., Patel, P., Scott, A., & Duran, X. (2002). Measuring third stream activities. Final report to the Russell Group of Universities. Brighton: SPRU, University of Sussex.

Nederhof, A.J. (2006). Bibliometric monitoring of research performance in the Social Sciences and the Humanities: A Review. Scientometrics, 66(1), 81–100.

Noyons, C. M. (2005). Science Maps Within a Science Policy Context. In H. F. Moed, W. Glänzel, & U. Schmoch (Eds.), Handbook of Quantitative Science and Technology Research (pp. 237–255). Springer Netherlands.

Olmos-Peñuela, J., Castro-Martínez, E., & D'Este, P. (2014). Knowledge transfer activities in social sciences and humanities: Explaining the interactions of research groups with non-academic agents. Research Policy, 43(4), 696-706.

Olmos-Peñuela, J., Molas-Gallart, J., & Castro-Martínez, E. (2014). Informal collaborations between social sciences and humanities researchers and non-academic partners. Science and Public Policy, 41(4), 493–506.

Rogers, J.D., & Bozeman, B. (2001). 'Knowledge Value Alliances': An Alternative to the R&D Project Focus in Evaluation. Science, Technology, & Human Values, 26(1), 23–55.

Spaapen, J., & van Drooge, L. (2011). Introducing 'productive interactions' in social impact assessment. Research Evaluation, 20(3), 211–8.

Upton, S., Vallance, P., & Goddard, J. (2014). From outcomes to process: evidence for a new approach to research impact assessment. Research Evaluation, 23(4), 352–65.

Watts, D.J., Dodds, P.S., & Newman, M.E.J. (2002). Identity and search in social networks. Science, 296(5571), 1302-1305.

AUTHORS

NICOLAS ROBINSON-GARCIA
INGENIO (CSIC-UPV), Universitat Politècnica de València
Valencia (Spain) Edif 8E 4º Camino de Vera s/n
46022 Valencia
Spain
elrobin@ingenio.upv.es
Phone number: +34 963 877 007
ext: 77048

THED N. VAN LEEUWEN
Centre for Science and Technology Studies (CWTS), Leiden University,
P.O. Box 905, 2300 AX Leiden
The Netherlands

ISMAEL RAFOLS
INGENIO (CSIC-UPV), Universitat Politècnica de València
Valencia (Spain) Edif 8E 4º Camino de Vera s/n
46022 Valencia
Spain

SPRU, University of Sussex
Brighton
United Kingdom

Observatoire des Sciences et Techniques, HCERES
Paris
France

EVALUATING TO VALORISE: THE SOCIETAL VALUE OF SSH RESEARCH AND THE ENRESSH COST ACTION

IOANA GALLERON, MICHAEL OCHSNER, JACK SPAAPEN AND GEOFFREY WILLIAMS

INTRODUCTION

In many areas of the SSH, evaluation has a bad press. It is often seen in antagonistic terms where policy makers impose ill-adapted procedures that distort SSH communication and dissemination traditions and which only serve in the handing out of funding resources, generally to the detriment of the SSH. At the same time, those SSH scholars who are against evaluation strongly believe in the value of their disciplines. They are also quite happy to evaluate their students and are firm believers in peer review. There is obviously a gap to be bridged, and the new COST action, ENRESSH, CA 15137, - the European Network for Research Evaluation in the Social Sciences and Humanities – sets out to do this.
This paper will set out to show the link between valorisation and societal impact of the SSH on the one hand, and research evaluation on the other hand. It discusses how ENRESSH will seek to involve the different stakeholders having a say in evaluation principles and processes. In a second part, this paper will present the provisional lessons learned from our initial survey about who's who in SSH research evaluation in Europe, and about how this is done.

SETTING THE SCENE

It is beyond doubt that the SSH make an important contribution to academia and society. There are many reports about the contributions of SSH research, and a fine showcase can be found on the REF-data base of case studies (http://impact.ref.ac.uk/CaseStudies/). But it is not only these specific SSH related results that count, it is also important to realise that many of the technical advancement evolving from STEM fields research, will only be accepted by the society when they are embedded in the social context and the concerns that are debated there about the impact of these technical innovations. This is where knowledge from SSH fields is often of vital importance. Whether it is about fracking, HIV vaccination campaigns, the migration crisis or global warming, in all these issues SSH expertise is much wanted to raise awareness, change attitudes, and alter policies.

Despite all this, many academics and policy makers fail to recognise the value of SSH research for the contributions they make in addressing the major societal challenges we are facing. Nevertheless, this situation seems to be slowly changing for the better. The EU framework program Horizon 2020 for example aims at giving SSH research a more prominent role in their programmes and policies. To date, this embedding process has had mitigated results (see European Commission, 2016) and, clearly, conditions need to be created for it to become more successful. The question is thus how we can monitor the valorisation of the SSH and, more importantly, how we can speed up this process. This is where the ENRESSH COST action comes in.

Enhancing the impact of SSH research – and of research in general - can only happen if three conditions are met simultaneously:

1. Academics' engagement with society's problems has to be made more explicit, both for the society but also for the scholars themselves;
2. Policy makers in academia have to become more aware of the value of SSH research for their own institutions, for other academic fields that engage with societal questions, and for society as a whole (notice for example that many universities owe their position in international rankings such as THE to a large extent to the SSH fields);
3. Evaluation policies at all levels have to become more comprehensive and inclusive for SSH as well as for all fields of the academic endeavour.

All three goals are equally important, but in this Action, the main focus is on the latter, evaluation and evaluation policy is at the centre of our activities. But we realise that the other two have to be part and parcel of the Action because without academics and policymakers realizing the vital importance of SSH research for the wellbeing and functioning of our democratic societies, no policy will likely succeed in changing things for the better.

THE ENRESSH COST ACTION

ENRESSH is a new dedicated COST action that sets out to tackle these three goals by bringing together stakeholders – institutional, policy and academic - from across Europe so as to propose new ways to look at research evaluation in the SSH. It aims to create a set of best practices that will not only lead to more constructive indicators, but actively valorise the research of the evaluates.

The main thrust of ENRESSH will be valorising SSH research. This calls for deepening our understanding of SSH research and their production and communication processes – both with regard to the scientific community and with societal partners. This means also studying the structural requirements and conditions necessary for the flowing of SSH knowledge and expertise towards other disciplines and society at large. Together, this will lay the foundation for an innovative approach to evaluation in which the contribution of the SSH to academia and society will be more adequately valued. With this, we will respond to a pressing need expressed by all parties involved in the exercise (i. e. evaluators, policy-makers as beneficiary of the evaluation, and last but not least scholars themselves).

COMPREHENSIVE EVALUATION IN A STAKEHOLDER CONTEXT

Clearly, one of the keys for a successful response is a productive co-operation between the various stakeholders, and this does not happen naturally. Firstly, because we are dealing with a variegated context in which many different stakeholders operate, each with their own expertise and interests; they are often not used to collaborate and as a rule have little spare time for talking to each other. Secondly, the SSH are known to be scientific fields with low degrees of functional and scientific dependence (Whitley, 1984), an affirmation more or less valid but with variations from one discipline to another and from one country to another. In the Netherlands, for example, many disciplines collaborate through national or at least supra-institutional plans (called discipline plans or research agendas), but in other countries this is not the case. Furthermore, as revealed by a recent analysis of the publication habits of scholars in the humanities, conducted on a sample of outputs by more than 300 Italian scholars over a ten-year period (Galleron and Basso, forthcoming), very different patterns of communication and dissemination exist, even between kin disciplines. Beyond the general picture one can get when looking at the whole, this discloses, unsurprisingly, the importance of books in many SSH fields. And the evaluation of book publications is under-developed compared to the evaluation of articles in scientific journals.

The variety and fragmentation of SSH fields may be a hindrance to coordination and collaboration, but at the same time it is important to realise that the diversity of this sector of knowledge is also its wealth and strength, in particular in relation to a society that is characterised by an ever accelerating pace. The involvement of scholars and stakeholders in the design of new evaluation protocols is therefore both necessary and complicated, and the challenge is then to come up with frameworks that do justice to both the disciplinary variation and the policy demands for frameworks that allow to some form of comparison necessary for policy and funding decisions.

EUROPEAN DIFFERENCES

It is important to realise that evaluation of and policy-making in the SSH are organised very differently from one European country to another, and even from one discipline to another. Furthermore, countries face different challenges in academia; thus, evaluations serve diverse needs. In some countries, SSH benefit from "special tracks" and specific protocols; in others, they are treated on the same basis as the "hard" sciences – which in some cases translates by "not treated" at all. Dedicated evaluation agencies intervene in certain cases, while ad hoc panels are put together in others. From a more qualitative perspective, expectations are clearly not the same towards the SSH, with some countries more interested in their societal relevance and impact, while others looking at the field from the point of view of its academic excellence, sometimes very narrowly defined as "capacity to publish in high impact factor journals".

ENRESSH sets out to tackle the problems related to finding innovative evaluation approaches that are able to cope with both types of diversity: that of the scientific variation in SSH research and of the multiplicity of societal demand and context. Its main aims are:

1. To improve evaluation procedures in order to take into account the diversity and the wealth of SSH research;
2. To make a robust case for the ways in which the SSH add value to the society;
3. To help SSH scholars better appropriate their research agenda and overcome fragmentation, while still leaving place for individual and specialized research typical for these disciplines.
4. To open up SSH research for interaction with societal stakeholders.

To achieve these results, ENRESSH brings together researchers from 29 European and cooperating states, whose perspectives on tools and methods can help tackling the complex problems of SSH evaluation. These researchers come from many different fields: sociologists and historians of science, bibliometricians, specialists in political sciences, as well as philosophers, cultural studies specialists, librarians and linguists. Together they will be able to oversee the developments in all these countries and in most SSH fields with regard to evaluation in the context of societal demand.

THE ACTION IS ORGANISED IN FOUR WORKING GROUPS:

1. **WG 1. Conceptual frameworks for SSH research evaluation:** The objective of this working group is to further develop our understanding of the SSH knowledge production processes and strategies, as a basis for developing evaluation procedures that adequately reflect the research practices, goals and aims of the SSH scholars. The working group will tackle the dialectic issues of the potentials and drawbacks of (a) metric approaches and peer review; (b) international exchange and cooperation and the local rootedness of SSH; and (c) the need for interdisciplinary exchange and disciplinary expertise.
2. **WG 2. Societal impact and relevance of the SSH research:** The objective of this working group is to analyse the engagement of SSH researchers with societal challenges, the ways in which interaction takes pace in non-academic partnerships and environments of SSH research, their diversity and their experiences with collaboration.
3. **WG 3. Databases and uses of data for understanding SSH research:** The main objective of this working group is to reflect upon the standardization and the interoperability of current research information systems dedicated to the SSH research outcomes.
4. **WG4. Dissemination:** The objective of this working group is on the one hand to ensure a maximum visibility of the Action through outreach activities among SSH researchers and specialists in research evaluation and among political, societal or economic stakeholders, and on the other hand this workgroup

is focussing on processes of engagement and communication between researchers and stakeholders.

A PAN-EUROPEAN PERSPECTIVE

ENRESSH takes a pan-European perspective. Analysing evaluation practices in the SSH in the European countries, we will describe and systematise the characteristics of various evaluation procedures in use and the ways policy makers and stakeholders are involved in designing and implementing these evaluations. We will synthetize the current strands of research on SSH evaluation, as well as the different philosophies (or doctrines) informing the exercise in the countries participating in the network.

A survey among the Management Committee members of the Action showed that there is a wealth of different evaluation procedures in Europe. The survey was conducted in the weeks before the kick-off meeting of the Action, i.e. in March 2016, and 45 representatives from 23 countries participated in the survey, i.e. one to five respondents per country. The results show that in a majority of the countries there is a national evaluation procedure in place. In slightly less countries, there is a national database of (SSH) publications at hand. In slightly more than half of the countries, SSH research is evaluated by at least some SSH-specific measures. However, the results also showed that existing typologies of evaluation procedures (Coryn et al., 2007; Hicks, 2010; 2012; Martin & Geuna, 2001; 2003; von Tunzelmann & Mbula, 1999) do not suffice to capture the diverse landscape of European research evaluation. First, representatives of the same country very often disagreed on many questions describing the evaluation system. Second, the "other" option was used quite often for many questions. A further interesting finding is the high number of respondents stating that in their country, there is officially no link between evaluations and funding but, in practice, evaluation results are used for funding decisions anyway.

We conclude from these preliminary results that a) definitions have to be made explicit (e.g. what exactly is performance based funding), b) typologies have to be expanded to be able to capture the diverse landscape of European evaluation procedures, c) awareness of the evaluation procedures in place at different levels has to be raised, and d) adequate evaluation procedures for the SSH have to be identified, optimised and promoted.

CONCLUSION

The problems facing the SSH and evaluation are not new. What is new is that they are finally being tackled across Europe in a concerted endeavour. ENRESSH brings together a community that has been studying evaluation phenomena for years, and offers the opportunity for others to join in the debate. ENRESSH has four years to carry forward its mission, and the foundations are already strong.

REFERENCES

Coryn, C. L. S., Hattie, J. A., Scriven, M., & Hartmann, D. J. (2007). Models and Mechanisms for Evaluating Government-Funded Research: An International Comparison. American Journal of Evaluation, 28(4), 437–457. http://doi.org/10.1177/1098214007308290

European Commission. (2016). Horizon 2020. Monitoring Report 2014. Brussels: European Commission. Retrieved from https://ec.europa.eu/programmes/horizon2020/en/news/horizon-2020-monitoring-report-2014

Galleron, Ioana, Basso, Antonella (forthcoming). "Publication habits and distribution of the output of Italian scholars from CUN areas 10 and 11"

Geuna, A., & Martin, B. R. (2001). University Research Evaluation and Funding: An International Comparison. SPRU Electronic Working Paper Series (Vol. 71).

Geuna, A., & Martin, B. R. (2003). University Research Evaluation and Funding: An International Comparison. Minerva, 41(4), 277–304. http://doi.org/10.1023/B:MINE.0000005155.70870.bd

Hicks, D. (2010). Overview of models of performance-based research funding systems. In Performance-based Funding for Public Research in Tertiary Education Institutions (pp. 23–52). OECD Publishing. http://doi.org/10.1787/9789264094611-4-en

Hicks, D. (2012). Performance-based university research funding systems. Research Policy, 41(2), 251–261. http://doi.org/10.1016/j.respol.2011.09.007

Tunzelmann, Von, N., & Mbula, E. K. (2003). Changes in research assessment practices in other countries since 1999: final report.

Whitley, Richard. (1984) The Intellectual and Social Organisation of the Sciences. Toronto: Oxford University Press.

AUTHORS

IOANA GALLERON
Contact author
Université Grenoble-Alpes (UFR LLSHS)
4, rue Jean Zay
56321 LORIENT CEDEX
FRANCE
Phone number: 00.33.6.52.63.36.22
galleron@evalhum.eu

MICHAEL OCHSNER
ETH Zürich
Switzerland
ochsner@evalhum.eu

JACK SPAAPEN
Royal Academy of Sciences
the Netherlands
jack.spaapen@knaw.nl

GEOFFREY WILLIAMS
Université Grenoble Alpes
France
williams@evalhum.eu

THE COMMUNITY INNOVATION SURVEY AND THE INNOVATION PERFORMANCE OF ENTERPRISES FUNDED BY THE EU'S FRAMEWORK PROGRAMMES.

LESSONS FOR THE EVALUATION OF HORIZON 2020'S ECONOMIC IMPACTS

JESÚS ALQUÉZAR SABADIE AND CLAIRE KWIATKOWSKI

1. BACKGROUND: INNOVATION AND IMPACTS IN THE EU'S R&D FRAMEWORK PROGRAMMES

Innovation is one of the key objectives of the ongoing European Union's Framework Programme for Research and Innovation, Horizon 2020 (2014-2020). It gained a special importance as a response the global economic crisis that started in 2007[1]. At that time, the Seventh Framework Programme of the European Community for research, technological development and demonstration activities (FP7, 2007-2013)[2] was mainly designed to implement the European Research Area (ERA). Its role in terms of innovation was to complement specific EU funding schemes, such as the Competitiveness and Innovation Programme (CIP)[3].

The political measures to tackle the economic and financial crisis such as the Barroso's Economic Recovery Plan[4], the Europe 2020 Strategy for a "smart, sustainable and inclusive growth"[5] and the Innovation Union flagship initiative positioned Research and Innovation (R&I) as an essential element to overcome the crisis, while addressing global societal challenges.

This historical background explains the strong focus of the current Framework Programme, Horizon 2020, on innovation. Indeed, the Commission sometimes presents Horizon 2020 as "the financial instrument implementing the Innovation Union, a Europe 2020 flagship initiative aimed at securing Europe's global competitiveness"[6].

Assessing the impact of R&I is also particularly important in Horizon 2020, which regulation contains many references like "achieving maximum impact", "achieving the greatest possible impact" or "maximise impact". The Horizon 2020 legal base includes a set of "performance indicators", including some few related to innovation: patents, share of participating firms introducing innovations new to the company or the market, or prototypes and testing activities.

There is a constant need to improve the evaluation and the monitoring system of the Framework Programmes. The High Level Expert Group for the Ex Post Evaluation of FP7 stated that „evaluation activities have been considered as routine activities in recent years (…). Considering that the Framework Programmes have consistently been the third largest budget of the European Union, a strategic and professional monitoring and evaluation system is required to increase transparency and serve as a comprehensive and trusted source of evidence-based decision making"[7]. This diagnosis is not new for the

1 What follows is based on Connolly et al. (2014) Ex-Post Evaluation of FP7, Cooperation Programme Theme: "Environment (including Climate Change). Report to the European Commission. Available at: http://bookshop.europa.eu/en/ex-post-evaluation-of-fp7-cooperation-programme-theme-environment-including-climate-change--pbKI0614224/

2 Note the different terminology used in the official FP7 and Horizon 2020 names. FP7 was about "research, technological development and demonstration" while Horizon 2020 focuses on "research and innovation". Innovation is for the first time explicitly mentioned in the name of the programme.

3 Decision 1982/2006 of the European Parliament and the Council of 18 December 2006 concerning the Seventh Framework Programme of the European Community for research, technological development and demonstration activities (2007-2013), Preamble, recital 22.

4 Communication from the European Commission (2008) A European Economic Recovery Plan, COM (2008)800 final.

5 Communication from the European Commission (2010) Europe 2020: A strategy for smart, sustainable and inclusive growth, COM (2010)2020 final.

6 See the Horizon 2020 website, at: https://ec.europa.eu/programmes/horizon2020/en/what-horizon-2020

7 Martinuzzi, A. et al. (2015) Commitment and Coherence, essential ingredients for success in science and innovation. Ex Post Evaluation of the 7th Framework Programme (2007-2013), p. 9.At: http://ec.europa.eu/research/evaluations/pdf/fp7_final_evaluation_expert_group_report.pdf#view=fit&pagemode=none

Commission services. Other previous evaluation exercises reached similar conclusions[8].

The current Horizon 2020 management is confronted to a paradox. Innovation is one of the core issues that the programme should address, but it remains weakly monitored. The Commission has not developed any specific tool to follow-up systematically and comprehensively the innovation results of projects and their impacts yet. At the eve of the Interim Evaluation of Horizon 2020, due on December 2017, basic information on innovation outputs and outcomes, like Technology Readiness Level (TRL) attained, barriers encountered to commercialise or exploit results, health/energy/resource efficiency/climate impacts of innovations (e.g. reduction of emissions, saving of energy or raw materials), commercialisation data or further investments committed is still missing.

How can the Commission assess the economic impacts of Horizon 2020, when the political priorities of President Juncker focus on growth, jobs and investment[9]? Politically, EU-funded R&I must demonstrate its impact and contribution to such economic goals, in a time when austerity measures strongly affect research funding in several European countries.

2. THE COMMUNITY INNOVATION SURVEY: A SOURCE TO ASSESS THE INNOVATION IMPACTS OF EU RESEARCH AND INNOVATION WORK PROGRAMMES.

It is unlikely that the Horizon 2020 projects' reporting will be improved and extended. The Commission considers that this would be contrary to simplification, a principle that guides the whole programme implementation. In its response to the High Level Expert Group for the Ex Post Evaluation of FP7 recommendations, the European Commission commits to "establish data links with external databases to complete and improve the quality of data sets"[10]. Alternative solutions and data sources must be therefore explored and used to assess the innovation results and impacts of funded projects.

Eurostat launched its first Community Innovation Survey (CIS) in 1992. This bi-annual large scale survey provides harmonised data on enterprises' innovation activities and results by sector, size of company, type of innovation and the various stages of the innovation process: objectives, sources of information, investments, public funding, etc.[11]

The CIS questionnaire includes an item that refers to funding from the Framework Programmes[12]:

During the three years 2010 to 2012, did your enterprise receive any public financial support for innovation activities from the following levels of government? (Include financial support via tax credits or deductions, grants, subsidised loans, and loan guarantees. Exclude research and other innovation activities conducted entirely for the public sector under contract).

> *Local or regional authorities – Yes/No*
>
> *Central government (including central government agencies or ministries) – Yes/No*
>
> *The European Union (EU) – Yes/No*
>
> *If yes, did your enterprise participate in the EU 7th Framework Programme for Research and Technical Development? – Yes/No*

The last question ("FUNRTD" variable) is very relevant for evaluating FP7 outcomes and impacts. It allows to identify in an aggregated way, within the CIS respondents, enterprises that received FP7 support. It makes possible to perform a counter-factual analysis in order to compare innovation results of companies that received EU funding with those that did not. The CIS data also permit to understand which factors and barriers influence innovation outcomes, both for EU funded enterprises and overall.

The CIS have been rarely used to evaluate the EU Work Programmes despite the need to obtain more data on innovation results and impacts than those coming from R&I projects' reporting. In 2009, a PRO-INNO report combined the analysis of CIS with ad-hoc surveys and case studies to conclude that "the Framework Programme attracts the highly innovative companies and research institutions in Europe". The participants were more R&D intensive, more networked and more internationalised than the average. They obtained higher returns on innovative sales[13]. The study referred to FP4 (1994-1998) and FP5 (1998-2002). Muldur et al.[14] reached similar conclusions in 2006, using the CIS 3 (1998-2000). In 2013, a Science-Metrix study on Small and Medium Enterprises (SMEs) innovation performance used the CIS 2010 to design and test an ad-hoc survey questionnaire[15] – never implemented.

Using the CIS for evaluation purposes presents some caveats that will be fully developed in the paper:

- Geographical coverage
- Timeframe
- Issues related to the questionnaire design

---

8 For instance, Connolly, N. et al. (2014) op.cit., p. 79, recommended the Commission to enhance its monitoring and follow-up system, especially in the areas of innovation and policy use of results. The authors of such assessment said that "the Commission [should be able] to identify innovative projects with potential societal impacts, as well as their strengths and weaknesses, to provide further support (if needed) and facilitate networking with complementary projects, and dissemination. For innovation issues, the monitoring system should rely on a set of smart indicators (…) and on insights from Project Officers".

9 See: https://ec.europa.eu/priorities/index_en

10 Communication from the European Commission(2016) Communication on the Response to the Report of the High Level Expert Group on the Ex Post Evaluation of the Seventh Framework Programme, COM (2016)5 final, p.9.

11 See: http://ec.europa.eu/eurostat/web/microdata/community-innovation-survey . The CIS 2012 methodology is explained in detail at: http://ec.europa.eu/eurostat/cache/metadata/en/inn_cis8_esms.htm

12 Question 5.3 in CIS 2012.

13 Fisher, R.; Polt, W. and Vonortas, N. (2009) The impact of publicly funded research on innovation: An analysis of European Framework Programmes For Research and Development. Luxembourg, OPOCE, pp.7-8.

14 Muldur, U. et al. (2006) A New Deal for and Effective European Research Policy. The design and impacts of the 7th Framework Programme. Dordrecht, Springer, p.129.

15 Hassan, E. et al. (2013) Testing Horizon 2020 performance indicator on SMEs involved in innovation projects under FP7. Brussels, Science-Metrix, produced for the European Commission, DG Research and Innovation

- Anonymisation of respondents[16]
- Problems of the eco-innovation module[17]

Despite these issues, the CIS remains a very relevant source of information to assess and analyse the innovation results of enterprises supported by the Framework Programmes. This is precisely the purpose of this paper. The authors use data from the CIS 2008, 2010 and 2012[18] and try to answer to the following questions: Do enterprises supported by EU perform better than the average? What are the factors and barriers to innovation, both for EU funded enterprises and overall? Are there significant differences by country, sector, size of enterprise, source of financing, etc.? What are the economic returns of exploited innovations? The analysis is based on FP7 funding but should also serve to demonstrate the need to further exploit the CIS data on a systematic basis for all the following EU Work Programmes. From the European Commission perspective, the final goal of the analysis is to extract concrete and operational lessons from FP7 which could be used for the Interim Evaluation of Horizon 2020.

The paper looks also at the results of the Eco-Innovation module proposed in the CIS 2008 and links its results with the general innovation trends. This is particularly relevant in the context of the Circular Economy strategy of the EU which should also be monitored[19]. The CIS 2014 will contain the next edition of the eco-innovation module.

3. PRELIMINARY RESULTS. DO INNOVATIVE ENTERPRISES SUPPORTED BY FP7 PERFORM BETTER? WHY?

A first relevant conclusion is that FP7-funded innovative enterprises perform significantly better than those not supported. Between 2006 and 2012, more than 70% percent of the firms that benefitted from FP7 funding introduced new products to the market, while others remained under 45%. The difference is less pronounced – but still very strong – when referring to products new to the firm only, while new to the market process innovations present lower figures (below 50% in all CIS 2008, 2010 and 2012) that anyway double those of enterprises not supported by FP7.

Of course, these significant correlations between FP7 participation and innovation performance do not necessarily means causality. It can be assumed that the Framework Programmes attract R&I-intensive organisations, which are expected to be more innovative than the average. FP7 was a programme focused on excellence, with a rather low success rate (18.7%)[20]. Beneficiaries needed to demonstrate very strong capacities to be selected by independent evaluators. Indeed, amongst the main FP7 beneficiaries we find the biggest European R&I organisations[21]. FP7 was likely to attract the most R&I intensive enterprises, which in turn improve their capacities thanks to collaboration in R&I at international level. The logic is likely to be circular, not linear.

The main differences between FP7 funded firms and enterprises not supported by the EU's Framework Programme appear for new to the market product and services innovations, and for new to the market processes. This indicates that FP7 led primarily to the development and implementation of novel products, services and processes and not to replicate or improve those that were already in the market. The opposite could be considered as a failure for a R&I programme of this magnitude.

However, those results hide differences by sector, size of enterprise and country. Not all kinds of enterprises that participate in FP7 obtain similar results. This is analysed in detail in the paper, with the scope of extracting conclusions for the Interim Evaluation of Horizon 2020.

AUTHORS

JESÚS ALQUÉZAR SABADIE
Contact author
European Commission
DG RTD-I1
CDMA 3/175
1049 Brussels
Belgium
Phone Number: 00 32 2 29 51167
jesus-maria.alquezar-sabadie@ec.europa.eu

CLAIRE KWIATKOWSKI
European Commission, DG Research and Innovation.

16 Mazzanti, M. et al. (2016) "Firm surveys relating Environmental Policies, Environmental Performance and Innovation.Design challenges and insights from empirical application", OECD Environment Working Papers No. 103. Paris, OECD.
17 Ibid, p. 24.
18 Their sample size and geographical coverage is presented in Annex 1.
19 Communication from the European Commission (2015) Closing the loop – An European action plan for the Circular Economy, COM(2015) 614 final
20 Source: CORDA database. The four main FP7 specific programmes ("Cooperation", "Ideas", "Capacities" and "People") plus Euratom received 135,799 proposals and only 25,363 where selected for funding.
21 See the main R&D investors in Europe at the annual EU Industrial Scoreboard reportsprepared by the European Commission, Joint Research Centre, at: http://iri.jrc.ec.europa.eu/scoreboard.html

SMASH WORKING PAPER ENVIRONMENTAL REGULATION AND ECO-INNOVATION: INSIGHTS FROM DIFFUSION OF INNOVATIONS THEORY

ABDELFETEH BITAT

1. EXTENDED ABSTRACT

Since it was first published in 1962 by Everett M. Rogers, the diffusion of innovations theory has been the subject of numerous applications in various fields. In his theory, E.M. Rogers explains how ideas spread through the process of adoption of innovations (Rogers, 2010). The applications of this theory went beyond its original domain. In fact, while Rogers (2010) has introduced his theory by explaining how technologies and best practices are adopted and spread among farmers, the methodology proposed by the disusion of innovations theory has been used in medical sciences (Greenhalgh et al., 2004), communication networks (Valente, 2005), marketing (Mahajan et al., 1990) or environmental innovation (Kern et al., 2005; Beise & Rennings, 2005; Lanjouw & Mody, 1996; Schwarz & Ernst, 2009) which we do as well in this paper. In 1991, Michael E. Porter published a short, yet controversial, article where he explained that stricter environmental regulation could, actually, improve business competitiveness through environmental innovation (Porter, 1991). This claim would later be known as the "Porter Hypothesis". It goes without saying that such a claim from an influential Harvard professor created a turmoil in the scientific, political and business community alike. Following this line of thought, the research in this paper is centered around the Porter Hypothesis. However, this paper limits itself to investigating the relationship between environmental regulation and environmental innovation, which is also known as the "weak" Porter Hypothesis (Mohnen & Van Leeuwen, 2015). Thus, the current paper is an addition to the scientific literature on the subject of the relationship between environmental regulation and environmental innovation based on the diffusion of innovations theory. Three policy alternatives, namely legally binding instruments, financial & market incentives, and self-regulation are compared in order to answer the following research question: which policy is more inclined to foster eco-innovation? To do so, this paper considers three theoretical approaches: the neoclassical, the evolutionary and the induced innovation theory. The relationship between environmental regulation and environmental innovation is tested using a German firm-based panel data collected by the Centre for European Economic Research in Mannheim (ZEW[1]) which will allow us to study the innovative behaviour of businesses over time. As noted by Jaffe & Palmer (1997) and more recently by Lanoie et al. (2008), the lack of dynamics is one of the recurrent shortcomings in testing the Porter Hypothesis. Accordingly, a dynamic count data model is compared to a static model in order to estimate the propensity of firms to innovate in response to a set of initiating factors for environmental innovation, namely the fulfilment of existing legal requirements, expectations towards future legal requirements, public funding, demand for environmental innovations and self-commitment (cf. figure ??). In addition, we control for research and development intensity and the size of the company. These factors are suspected to be responsible for an important omitted-variable bias causing model misspecification (Griliches, 1979; McWilliams & Siegel, 2000). We also control for the region (eastern/western Germany) and the potential industry bias by using 23 sectoral dummies, and filter for companies that account for their environmental impact (Wagner, 2010; Busch & Hoffmann, 2011). The remainder of this paper is divided into five sections, a review of the relevant literature on the diffusion of innovations theory and the relationship between environmental regulation and environmental innovation, followed by a formulation of the hypotheses to be tested, the methodology used for the empirical model, the results and a discussion of these results.

2. DISCUSSION

The empirical results agree with the hypotheses formulated and the findings of previous research. Furthermore, they allow to shed some light on an important question. If environmental regulation is indeed necessary in order to trigger environmental innovation, how should it be designed?

2.1. NEOCLASSICAL APPROACH

To answer this central question based on the neoclassical approach, we compare different policy alternatives, namely legally binding instruments, financial and market incentives and self-regulation.

2.1.1. LEGALLY BINDING INSTRUMENTS

When studying the first alternative, two initiating factors were analysed: the fulfilment of existing legal standards and the expectation towards future legal requirements. Both theoretical and empirical evi-

1 ZEW stands for Zentrum für Europäische Wirtschaftsforschung.

dence point to the relative ineffectiveness of existing regulation obligation compared to long term performancebased regulation when the aim is to foster environmental innovation, dynamically. As a matter of fact, the former option limits the choice of businesses in term of the technology used to meet the regulatory requirements. On the other hand, performance-based regulation gives businesses the freedom to choose the best technology, and at the same time encourages them to find new, more eÐcient and effective techniques to meet the long-term regulatory objectives. Besides, when standards are based on a specific technology, they not only encourage end-of-pipe solutions, but may also discourage innovative behaviour due to the regulatory uncertainty inherent to such regulation. That is to say, businesses may refrain from innovating in apprehension of a rise of the regulatory standard. In contrast, performance-based regulation set long-term objectives that are systematically reviewed over a known time-horizon, thus it creates a market for environmental innovation and encourages businesses to find better ways to meet the regulatory objective. Nevertheless, for elected policy-makers, the choice of standard-based environmental regulation over performance-based regulation is motivated by two arguments. The outcome of the latter is less certain and requires longer periods than the former, in addition to diÐculty of setting the long-term objectives with the right balance between environmental protection and economic growth. In fact, the objectives should be both ambitious and realistic, otherwise they will either fall short of environmental protection, or will hamper economic growth. An other argument in favour of legally binding instruments is intrinsically linked to the nature of environmental innovation with a distinction between end-of-pipe innovation and other forms of innovation. In fact, the use of end-of-pipe solution might be necessary awaiting a more radical solution.

2.1.2. INCENTIVES

The second policy alternative is financial and market incentives. When studying this alternative, two initiating factors were analysed: public funding and demand for green products. The theoretical arguments could not provide a clear-cut on the effectiveness of financial incentives to foster environmental innovation. Neither did the empirical results. In fact, we show that these instruments are positively associated with environmental innovation only when they are forward looking such as the expectation towards a market demand for green innovation. That being said, it is important to distinguish between price and quantitybased instruments on the one hand, and information-based instruments on the other hand. Although it is necessary to correct market failures inherent to eco-innovation, such as the spillover effect, the former alternative may delay eco-innovation if the design of a subsidy is flawed, due to regulatory capture where special interests affect regulatory intervention in setting R&D subsidies for instance (Dal B´o, 2006). In that case, subsidies may even lead to a perverse effect where businesses rely on end-of-pipe solutions, only to avoid any compliance penalties, because of the lack of an incentive and reward system for innovation beyond compliance. In comparison, information-based instruments rely on improving information flows in order to harness market forces with the aim of fostering eco-innovation. In fact, by educating both the consumer and the producer, policy-makers will create an environment where there is a demand, and thus a market, not only for green products but also for green innovations. These forces can then act freely under the market conditions where the choice of the best technology will be decided based on its effectiveness and effciency. The diffusion of such technologies will in turn improve its economic performances, thanks to scale, scope and learning effects. At the same time, the demand for green products will create a sound competitive environment for innovators racing to find the next standard-setting technology. Nonetheless, the limits of such policy is the uncertainty around the outcome and the time necessary to reach the intended results. It is also important to note that environmental and technology policy are more effective when the regulator should enable ecological modernization rather than controlling the process of transition.

2.1.3. SELF-REGULATION

Lastly, a third alternative is tested using self-commitment as an initiating factor for environmental innovation. As expected, the empirical results confirmed the theoretical arguments. Clearly, if left unregulated, businesses would not choose to eco-innovate. The decision is based on solid motives which are, unfortunately, not socially optimal. To put it differently, with no regulatory constraints, businesses would not have to in-ternalise the cost of their negative externalities when harming the environment. Additionally, in the event that they decide to eco-innovate, businesses will refrain from doing so continuously for the simple reason that while the whole society benefits from eco-innovation, the sole bearer of the cost is the innovator, not to mention the fact that the technology can then be copied, thus stripping it of its competitive advantage. Finally, the reason no, or little, eco-innovation should be expected without stringent environmental regulation is, partly, because other investment alternatives are, usually, more financially rewarding. In that respect, policy-makers should act to improve the financial attractiveness of investments in environmentally friendly technologies. Therefore, regulatory intervention is, indeed, the sine qua non of environmental innovation.

2.2. EVOLUTIONARY APPROACH

When exploring the evolutionary, we first came to the conclusion that self-regulation can indeed be effective if, and only if, coupled with the expectation of stringent future regulation and collective liability in case failure in meeting the objectives agreed upon. This situation has been illustrated in different cases such as the Dutch energy benchmarking program or the discussion between the European Commission and the Auto industry in Europe. Del R´ıo et al. (2010) refers to this approach as "combining the carrot and the stick".

2.3. INDUCED INNOVATION APPROACH

Lastly, exploring the induced innovation approach allowed us to acknowledge that while innovation is path dependent, the decision to eco-innovate can be stimulated by internal factors, such as cost savings, rather that external factors only. In that sense, national innovation systems need to be adapted in order to break path dependency on old, polluting technologies and stimulate technological change by enabling it rather than controlling all its aspects. To do so, a clear long term vision need to be shared by the stakeholders (economic, public and civil). However, these objectives need to be updated dynamically and systematically, otherwise too much certainty might inhibit eco-innovation.

CONCLUSION

The objective of this paper is to answer the following research question: which policy is more inclined to foster eco-innovation? To do so, we base our analysis on three theoretical approaches: neoclassical, evolutionary and induced innovation. However, the research on the relationship between policy and the diffusion of innovations is not limited to these three approaches. In fact, some researchers have based their studies on the actor-networks theory (Braun, 2008; Truffer & Coenen, 2012), the systemic approach (Edquist, 1999) or the practice-based approach (Mele & Russo-Spena, 2015). These approaches explore different aspects of eco-innovation such as the role of the position in the network, the role of national innovation systems and institutional factors as well as the non-linear and dynamic nature of innovation. These limitations should be addressed in further research. Similarly, notwithstanding the fact that we control for the sector and the size of businesses, an investigation comparing different sectors or business sizes should allow to draw more practical policy recommendations. By contrast, combining data on different countries over a longer time frame should allow to draw conclusions with more perspective. Nonetheless, the results of this paper allow us to draw the following policy recommendations: conventional regulatory tools, namely legally binding instruments are not effective for triggering innovative behaviour at the firm level while market incentives have a positive effect on the diffusion of innovations. Moreover, there is a market inertia justifying regulatory intervention in order to break path dependency with innovative policy instruments that create a sound and dynamic environment for eco-innovation. In fact, environmental policy should not be "slow, cumbersome, expensive, uncoordinated and uncertain" (Palmer et al., 1992, p. 259), but rather be "proactive, ambitious, open, flexible and knowledge oriented" (Del Rio et al., 2010, p. 547) arising from dialogue and consensus. Similarly, the objective of the interaction of technology policy and environmental policy is by no mean to penalize polluting businesses but rather to lift the barriers to ecoinnovation allowing the passage to a more sustainable economy (Del Rio et al., 2010), what Huber (2000) qualifies of "ecological modernization".

AUTHOR

ABDELFETEH BITAT
Universite Saint-Louis Bruxelles
Faculty of Economics, Social and Political Sciences and Communication
Boulevard du jardin botanique 43
Brussels 1000
Belgium
Tel.: +32 485 806 130
Fax: +32 2 211 79 97
abdelfeteh.bitat@usaintlouis.be

INTELLECTUAL PROPERTY RIGHTS AND TECHNOLOGICAL INNOVATION: CASE STUDY OF RENEWABLE ENERGY ADOPTION

OLUWASOLA EMMANUEL OMOJU

1. BACKGROUND

The main objective of intellectual property rights (IPRs) protection is to foster innovation and economic development. However, its impact has been a subject of intense controversy since the ratification of the World Trade Organisation Agreement on Trade Related Aspects of Intellectual Property Rights (TRIPS) in 1995. While a school of thought argues that IPR protection enhances economic growth and development (Hu and Png, 2010; Branstetter, Foley and Saggi, 2010), another school argues that it hampers it (Brenner-Beck, 1992; Chaudhuri, Goldberg and Jia, 2006). The advocates of IPR posit that strong IPR will promote economic growth and development by increasing the incentives for innovation and encouraging foreign direct investment (FDI). On the other hand, the antagonists of IPR opine that it enhances the monopoly powers of innovators and limits the ability of developing countries to access importation innovations.

Among the potential areas where the impact of IPR protection is pronounced is clean technology transfer. Over the past few years, policy makers have been initiating policies to address increasing carbon emission, which is the primary cause of climate change. Global carbon emission increased from 22.7 billion metric tonnes in 1990 to 34.5 billion metric tonnes in 2012, indicating an increase of 52% (Oliver, Janssens-Maenhout, Muntean and Peters, 2013). One of the policy measures proposed by policy makers and other stakeholders in order to reduce energy-related carbon emission is innovation and massive deployment of renewable energy. While some developed countries have significantly advanced innovations in clean technologies and increased their shares of renewable energy in total energy supply, developing countries still generates only a negligible portion of energy use from renewable sources. According to Latif (2013), six OECD nations account for almost 80% of all patent applications in clean energy technologies. Given the differences in resource endowment, level of economic development and technological advancement among countries, there is need for technology transfer from technologically-advanced countries to technologically backward countries.

2. STATEMENT OF THE PROBLEM

There have been numerous studies on the economic impacts of IPR but these studies have concentrated on the impact of IPR on foreign direct investment (FDI), industrial development, health and economic growth (Bransteter and Saggi, 2009; Correa, 2002; Branstetter, Foley and Saggi, 2010). Surprisingly, studies on the impacts of IPR on clean technology transfer are rare. According to Latif (2012), technology transfer and innovation are crucial issues in the Rio+20 United Nations Conference on Sustainable Development discussions held in 2012, but IPR which are related to them are barely mentioned. This is in sharp contrast with the 1992 Rio Earth Summit, which included several provisions of IPR and green technologies. IPR is a major decisive factor in technology transfer, but empirical analysis of the impact of IPR on renewable energy adoption and sustainable development are very rare. Majority of available studies that have examined the effects of IPR from a clean energy technology and climate change angle are mainly from theoretical and legal perspectives (Barton, 2007; Tomlinson, Zorlu and Langley, 2008; International Centre for Trade and Sustainable Development, 2008; Kapur, 2011; IRENA, 2012; Latif, 2013). Among the few empirical studies, there is evidence of a positive association between IPR and technology transfer (Branstetter, Fisman and Foley, 2006; Nicholson, 2003; Ivus, 2008). But latter studies argue that the conclusion seem to hold only for large and middle-income emerging countries with substantial ability to adapt technologies (Maskus and Okediji, 2010). However, these studies examine overall technology and are not specific to renewable energy technologies.

3. CONTRIBUTION OF THE STUDY AND RESEARCH QUESTIONS

Therefore, there is need for an empirical study of the relationship between IPR and renewable energy adoption to provide a basis for more encompassing and evidence-based discussions on the subject. This is the major contribution of this paper to the literature on IPR and technology transfer. Specifically, this study investigates the impact of IPR protection on the adoption of renewable energy in a panel of developed and developing countries. Unlike previous studies that adopt theoretical and legal approach, the present study uses econometrics methods to analyse the impact of intellectual property right policies on renewable energy technology adoption, after controlling for other determinants of renewable energy adoption. It also investigates whether the impact of IPR protection on renewable energy adoption is dependent on the local conditions of the country. The research questions this paper seeks to answer are: Does stronger IPR protection enhance or impede renewable energy adoption? Is the impact of IPR protection on renewable energy adoption dependent on local conditions? Answering these questions is the aim of the study.

4. DATA AND METHOD

Data for this study includes panel data of 102 countries from 1990-2010 with 5-year intervals. The data on IPR protection is from the Ginarte-Park index. The Ginarte-Park index developed by Ginarte and Park (1997) and Park (2008) ranks countries on the strength of intellectual property rights protection and includes data on 122 countries globally. The index ranges from 0 to 5, with 0 indicating the lowest level of IPR protection and 5 indicating the highest level of IPR protection. The index is calculated based on a survey conducted in these countries every five years. This index is preferred to indicate IPR protection in this study because it indicates the strength or degree of protection, and the data is available for a large number of countries over an extended period of time. Therefore, IPR (which is denoted by the Ginarte-Park index) is the variable of interest in the study. On the other hand, data on renewable energy adoption is obtained from the World Development Indicator (WDI) database of The World Bank. In this study, rather than focusing on the size or amount of renewable energy, the study analyses the share of renewable energy in total energy use. This is important because it is the share of renewable energy in total energy use (not the amount) that is important for decarbonisation of the energy systems and climate change mitigation. The share of renewable energy in total energy use is the dependent variable in this study. In order to specifically determine the impact of IPR protection on renewable energy adoption in this study, controlled variables are included in the model to account for other important drivers of renewable energy adoption. The control variables in this study include GDP per capita (denoting the level of development), trade openness, research capability and the square of IPR. The square of IPR is included to investigate the evidence of a non-linear relationship between IPR and renewable energy adoption. The estimation technique in this study is the fixed and random effect regression and the Hausman specification test.

5. RESULTS AND POLICY IMPLICATIONS

The results of the study show that stronger IPR protection undermines renewable energy adoption after controlling for other drivers of renewable energy. Economic development has a negative impact on the share of renewable energy in total energy use while trade openness has a positive impact. The evidence for a non-linear relationship between IPR protection and renewable energy adoption is weak as the coefficient is only significant at 10% level. The results further show that the impact of IPR protection on renewable energy adoption is partly dependent on the level of development and the research capability of a country. In other words, with the same level of IPR protection, higher economic development and research capabilities reduces the negative impact of IPR on renewable energy adoption. Following the results of this study, the paper suggests that policy makers should design IPR frameworks and regulations in such a way that it will not deter the transfer and adoption of key technologies like clean energy technologies. IPR framework should be designed in ways that enable access to critical technologies without discouraging or stifling innovation. As technology transfer mainly takes place through trade and foreign direct investment, it is suggested that countries adopt open trade and investment policies that encourages the flow of clean technologies. Furthermore, policies that support economic development and the research capabilities of countries should be promoted to minimise the negative impact of IPR on renewable energy adoption.

6. REFERENCES

Barton, J. H. (2007): “Intellectual property and access to clean energy technologies in developing countries: An analysis of solar photovoltaic, biofuels, and wind technologies”, International Center for Trade and Sustainable Development, Trade and Sustainable Energy Series, issue paper no. 2.

Branstetter, L. G., R. Fisman and C. F. Foley (2006): “Do stronger intellectual property rights increase international technology transfer? Empirical evidence from U.S. firm-level panel data”, Quarterly Journal of Economics, vol. 121, pp. 321–349.

Branstetter, L. and K. Saggi (2009): “Intellectual property rights, foreign direct investment, and industrial development”, National Bureau of Economic Research, working paper series no. 15393.

Branstetter, L., C. F. Foley and K. Saggi (2010): “Has the shift to stronger intellectual property rights promoted technology transfer, FDI, and industrial development?” WIPO Journal, vol. 2, pp. 93-98.

Brenner-Beck, D. (1992): “Do as I say, not as I did”, UCLA Pacific Basin Law Journal, vol. 11, pp. 84-118.

Chaudhuri, S., P. Goldberg and P. Jia (2006): “Estimating the effects of global patent protection in pharmaceuticals: A case study of quinolones in India”, American Economic Review, vol. 96, pp. 1477-1514.

Correa, C. M. (2002): “Public health and intellectual property rights”, Global Social Policy, vol. 2, pp. 261-278.

Ginarte, J. C. and W. G. Park (1997): “Determinants of patent rights: A cross-national study”, Research Policy, vol. 26, pp. 283-301.

Hu, A. G. Z. and I. P. L. Png (2010): “Patent rights and economic growth: Evidence from cross-country panels of manufacturing industries”, Paper presented at the WIPO Seminar Series on “The Economics of Intellectual Property”, Geneva, November, 2010.

International Centre for Trade and Sustainable Development (2008): “Climate change, technology transfer and intellectual property rights”, Paper presented at the Trade and Climate Change Seminar co-organized by the Government of Denmark, the German Marshall Fund of the United States and the International Institute for Sustainable Development, Copenhagen, Month, 2008.

IRENA (2012): “Assessment of intellectual property rights for promoting renewable energy”, Paper presented at the IRENA Roundtable Workshop, Bonn, October 2012.

Ivus, O. (2008): "Do stronger intellectual property rights raise high-tech exports to the developing world?" University of Calgary, Department of Economics, working paper no. 2008-27.

Kapur, K. (2011): "Climate change, intellectual property, and the scope of human rights obligations", Sustainable Development Law & Policy, vol. 11, pp. 58-63, 95-97.

Latif, A. A. (2012): "Intellectual property rights and green technologies from Rio to Rio: An impossible dialogue?" International Center for Trade and Sustainable Development, policy brief no. 14, July 2012.

Latif, A. A. (2013): "Intellectual property rights and clean energy technologies", Paper presented at the United Nations Workshop on Clean Energy Technologies, New York, May 2013.

Maskus, K. E. and R. L. Okediji (2010): "Intellectual property rights and international technology transfer to address climate change: Risks, opportunities and policy options", International Center for Trade and Sustainable Development, Intellectual Property and Sustainable Development Series, issue paper no. 32.

Nicholson, M. (2003): "Intellectual property rights, internalisation and technology transfer, US Federal Trade Commission, working paper.

Oliver, J. G. J., G., Janssens-Maenhout, M. Muntean and J. A. H. W. Peters (2013): "Trends in global CO2 emissions: 2013 report", PBL Netherlands Environmental Assessment Agency, The Hague.

Park, W. G. (2008): "International patent protection: 1960-2005." Research Policy, vol. 37, pp. 761-766.

Tomlinson, S., Zorlu, P. and Langley, C. (2008): Innovation and technology transfer: framework for a global climate deal, An E3G and Chatham House Report, London.

AUTHOR

OLUWASOLA EMMANUEL OMOJU
China Center for Energy Economics Research (CCEER)
Xiamen University, Xiamen, 361005,
China
+8618559654314
shollcy@yahoo.co.uk

NEW EVALUATION APPROACHES – DATA, METHODS, INDICATORS BEYOND THE ORDINARY - EXPERIMENTS IN RTI EVALUATIONS: INSPIRATIONS, SUCCESSES AND FAILURES

DOMENICA BLUNDI

1. INTRODUCTION

This paper aims to describe the experience of a Brazilian mining company in the development and implementation of R&D management processes, focusing its project assessment methodology (ex-ante evaluation; mid-term evaluation; and ex-post evaluation). First, I present the arguments of the theoretical school Variety of Capitalism[1] (or simply VoC) in order to base the case presented here. Following, I mention the role of „institutions" in the analysis of innovative processes, presenting more specifically the latest change in the institutional framework of Brazilian Science Technology & Innovation (ST&I). Then, I will make a brief account of the company's case to, after, describe the methodology of R&D management and evaluation adopted, with some highlights and lessons learned throughout its implementation. At the end, I share some considerations and points of attention in order to contribute to the discussions involving new forms of R&D evaluation in specific institutional frameworks and, therefore, new variables that might compose the national ST&I policies.

2. CAPITALIST ECONOMIES AND ITS GOVERNANCE FRAMEWORK

The theoretical school Variety of Capitalism believes that companies are relational actors, whose main objective is to develop productive capacity and competence through interaction with the context from which they are part. According to this approach, the agents of the capitalist system promote the link between institutional changes and adaptive adjustments necessary for survival in the new order (Hall; Soskice, 2001, p.6). Thus, the organization of capitalist production should be considered on possible interactions between the productive sectors and other institutions. Here, the object of analysis are the adaptive processes of institutional actors in changing situations and their ability (or non-ability) of coordination.

From this assumption, Soskice (1999) stipulated two „ideal types" of capitalist economies using for this, the concept of „productive regimes" (how the production of certain sector is organized according to the context in which it operates): the „coordinated market economies" (CMEs) and „liberal market economies" (LMEs)[2]. In CMEs, the main mechanism for coordination of economic actions are relations of non-market, whereas in LMEs, the market is the main reference. In other words, the former tend to more collaborative practices between stakeholders, while the latter, for smaller collaborative practices.

Table I: The patterns of productive regimes of capitalist economies and its governance framework (Soskice, 1999)

CMEs	LMEs
FINANCIAL SYSTEM	
Long-term financing	Short-term financing
INDUSTRIAL RELATIONS SYSTEM	
• Trade unions are incorporated • Wage bargaining between companies	• The unions are less embedded • Unilateral agreements by management
SYSTEM OF EDUCATION AND TRAINING	

1 Hall, P.; Soskice, D. (2001). Varieties of capitalism. The institutional foundations of comparative advantage. New York: Oxford University Press.

2 CMEs would be represented by the Scandinavian countries, Germany and Switzerland, and on a different level, Japan and South Korea; LMEs and the Anglo-Saxon economies and Ireland. Countries like France and Italy would be variants of these major patterns. Not included are NICs, the economies of Eastern Europe and the less developed economies of Europe, such as Portugal, Spain, Greece and Turkey.

• Long-term training based on the vocation of the young • Conjunction within university system and internal R&D	• Emphasis on more general education • Lack of investment in long-term of vocational training • Little coordination within university system and internal R&D
INTER-COMPANY SYSTEM	
• Industrial coordination • Technology transference • Standardization of conduct	• Industrial competition • Competition for technology • Heterogeneity of conducts

Without minimizing the VoC assumptions, Lazonick (2007) - while conducting a comparative observation of innovation enterprises in capitalist economies - states that "to understand 'varieties of capitalism', one must begin with an analysis of the role of innovative enterprise – the quintessential capitalist institution – in the development of the economy"[3]. According to him, one should consider the analysis of "varieties of innovation enterprises" in order to provide new paths to be explored within specific domestic contexts (or "social conditions")[4]. This, in itself, justifies the analysis of its relationship with the institutional framework which they are embedded.

For the purposes of this work, and based on the brief theoretical framework presented so far, I will limit myself to three assumptions over the next few sessions, namely: (a) the strategies carried out by innovative enterprises should be analyzed with a view to 'productive regime' of which they are part, which may involve more or less coordination between the actors involved; (b) innovation is sine qua non for the survival of capitalist enterprises and, for this, is necessary that they develop new management models to deal with changes in institutional innovation contexts; and (c) given that innovative companies are the actor par excellence of capitalist economies, the investigation of their behavior is relevant to provide subsidies for a better analysis of ST&I government policies.

3. THE ROLE OF INSTITUTIONS AND THE BRAZILIAN POLICY OF ST&I: LIMITS AND SCOPE

According to Lazonick (2007, p.6), "nations differ historically in relation to their institutions". At some point along its trajectory, such institutions stimulate or undertake activities developed by economic agents; while, in turn, some institutional aspects become incorporated into the modus operandi of these agents[5]. Here, the concept of "institution" considers both the formal rules (laws, constitutions and property rights) as well as informal rules (sanctions, taboos, customs, traditions and codes of conduct). In this sense, the formal and informal rules, together, define the structures that support and mobilize social relations, specifically economic relations (North 1998, p. 248).

In order to characterize the institutional environment in which the Brazilian innovative companies interact, will present the latest changes within the ST&I policy, known in Brazil as "New Legal Framework for Science, Technology and Innovation"[6].

A. THE NEW LEGAL FRAMEWORK FOR SCIENCE, TECHNOLOGY AND INNOVATION IN BRAZIL

Despite having a broad institutional framework of ST&I, Brazil cannot break some barriers with regard to the relationship between investment in ST&I and economic development of the country. As stated by the World Bank (2016), in 2013 were invested just over 1.24% of GDP in Brazilian ST&I, a proportion far behind countries like Israel (4.11%), Korea (4.29%), Finland (3.17%) and Japan (3.58%).

According to a technical study produced at the request of the Legislative Chamber of the Brazilian government[7], within the bounds indicated as limiting to the ST&I advance in Brazil would be "the isolation of the academy; excessive bureaucracy and lack of decentralization mechanisms". These aspects were raised when drafting the Project Law nº 2.177/11, at which representatives of the main Brazilian ST&I organizations were consulted. In this dialogue, we identified that the most important points to transform the ST&I in the country are the integration of the scientific community with the private sector and greater flexibility regarding the university-industry regulamentation (Nazareno, 2016, p.6).

Bearing in mind that innovation is an intrinsic practice of capitalist economy models, representing even a relevant condition for companies to survive, will present the following description of the experience adopted in a Brazilian mining company, with regard to the design of a management and evaluation of R&D projects, where he sought to establish collaborative practices, through which has attempted more coordination with the ST&I community, decreasing its distance.

4. A BRAZILIAN MINING COMPANY: BRIEF DESCRIPTION[8]

FIRM is a Brazilian company founded in 1942, present in six continents and a world leader in the production of iron ore pellets and nickel. Your investments also include logistics, steel and energy, as well as being committed to environmental responsibility and social activities.

Specifically in relation to technology and innovation activities, FIRM presents major challenges: has a complex and sophisticated logistics

3 Lazonick, W. (2007): ÐVarieties of Capitalism and Innovative EnterpriseÐ, in Lars Mjøset, Tommy H. Clausen (ed.) Capitalisms Compared (Comparative Social Research, Volume 24), Emerald Group Publishing Limited, p. 1.

4 Lazonick, W., idem, p.2.

5 Other authors also stressed the role of institutions as a relevant variable in the analysis about the performance of economic agents in different nations, namely: Williamson (1996); North (1998).

6 The "New Legal Framework for Science, Technology and Innovation" (Law nº 13.243, January, 11th, 2016) introduces important changes in the called Innovation Law (10.973/04), creating new institutional framework for Brazilian innovative ventures.

7 Nazareno, C. (2016). "As mudanças promovidas pela lei nº 13.243, de 11 de janeiro de 2016 (Novo Marco Legal de Ciência, Tecnologia e Inovação) e seus impactos no setor". Estudo Técnico. Câmara dos Deputados, Brasília, DF.

8 In order to maintain the integrity of the company presented, I will use the fictitious name "FIRM".

(about 5,000 km of railways; nine ports); deals with an intensive consumption of energy in its operations; generating hundreds of millions of tons of ore per year; search exploration and mineral processing technologies increasingly advanced; besides being committed to minimizing environmental impacts and health and safety in their endeavors.

Until the mid-2000s, FIRM maintained a closed innovation model, which prioritized R&D projects carried out internally by their own areas of business. His few external partners were triggered to execute specific projects, within a logic of providing research services. Moreover, their technology processes were decentralized and did not meet a specific governance. It can be said that his model was quite compatible with a little given production system to collaborative activities, close to the kind adopted by companies of LMEs, where the sharing of experiences and exchange of information are more timid and punctual.

In 2009, FIRM created a board responsible for establishing a R&D management model, in order to improve the coordination and integration of innovation processes. In addition to creating such processes, this board was also responsible for developing a research institute whose main objective would be to ensure a long-term look considering the future challenges of the company, while the business areas would be dedicated to the operational improvements, more immediate ones, which solutions do not require a high degree of innovation. One of the principles carried out by the new management was to increase and diversify the portfolio of the company's research partners, approaching the open innovation model and encouraging the integration with national and international ST&I community. It can be said that this event was an approach toward a more collaborative production system model, like the economic agents present in CMEs.

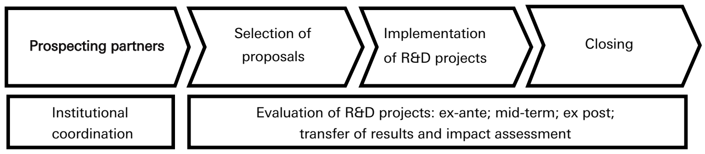
Let's see how this change has resulted in a methodology for the management and evaluation of R&D projects and what was the outcome for the company so far..

5. THE METHODOLOGY OF MANAGEMENT AND EVALUATION OF R&D PROJECTS: A COLLABORATIVE MODEL

Between 2010 and 2014, a working group was dedicated to develop, test and implement a methodology for the management and evaluation of R&D projects focusing on the establishment of partnerships with universities and research centers. Imbued with the premise that the company needed to expand and diversify its portfolio partners to perform R&D "extramural", this group sought to develop a methodological design that would guarantee the integration between the various actors involved. The proposal behind the model was: (a) have a more coordinated access to new knowledge in ST&I, acting more prospectively; (B) ensure a long-term look at the challenges; and (c) to seek cooperation with the ST&I community, to the detriment of specific partnerships, marked by a kind of relationship "customer service-provider".

To do this, it was created the following management and evaluation of R&D projects:

Figure I: R&D management and evaluation process



- Prospecting partners: this step can be conducted either by the Call of Proposals launch in conjunction with government agencies such as through partnerships with a specific university. In both forms, is invested in a strong work of institutional coordination, which seeks to complementarity of interests between stakeholders, away from the model order / contract research.
- Selection of proposals: at this stage, we seek to combine three dimensions of ex-ante evaluation: scientific, technological and business strategy. For each one, a group of assessors is triggered (including ad hoc evaluators), in order to ensure different perspectives on the displayed content and, therefore, ways to combine different interests. It would be a form of "collaborative evaluation" in order to achieve common gains for those involved: the company itself, its partner (e.g.: university or research group) and the ST&I policies, whereas only projects that contribute to the implementation of the local ST&I policy and the advancement of scientific knowledge are approved.
- Implementation of R&D projects: This step keeps the logic carried out in the "Selection of proposals", where the three dimensions of evaluation – scientific, technological and business strategy - go together. The mid-term evaluations of the projects are executed both through reports submitted, and through participation in workshops and technical visits, seeking the exchange of ideas and interaction between the experts of the company and the groups responsible for implementing the project within the institution. It must be said that, here, the latter are as purposeful as the contractor, as it is understood that the R&D project should be run through interactive actions, seeking a joint learning process.
- Closing and results internalization: upon completion of the project, we seek to evaluate the potential continuity of actions generated there. If the experience of the partnership was positive, generating common benefits, and if there is mutual interest in maintaining cooperation, starts new mechanisms for coordination between the business and the research institution. The proposal is to seek a long-term look, the understanding that the positive impacts of R&D projects should be achieved through institutional cooperation at the expense of relationships to provide timely research services and short-term.

Since this methodology has being adopted by FIRM to manage and evaluate its R&D portfolio, is possible to identify some relevant results:

Collaborative Model	
Before	**After**
Lack of R&D networks – many projects being developed by few institutions	Diversified, wider and better R&D partnerships: geographically and thematically
Personalized interaction between company and research group	Institutionalized interaction between company and research groups
R&D portfolio concentrated in incremental and short-term challenges	R&D portfolio more able to deal with disruptive and long-term challenges
Exclusive source of financial resources: the company itself	Partnerships with agencies and/ or research funding bodies: leveraging financial resourcers

6. FINAL CONSIDERATIONS

The R&D management and evaluation processes presented here has been tested by FIRM for five years. Its main assumption is to implement a non-restrictive process, where different "looks" can be complemented by encouraging the coordination of interests involved and generating common benefits.

Currently, FIRM has a portfolio that provides access to a wide universe of R&D projects. Through this portfolio, the company now has expanded its horizon, achieving a more active position in ST&I community. Since then, FIRM has been recognized as an actor that proposed amendments to the dynamics of university-industry relationship commonly practiced in Brazil, seeking to implement a differentiated model of interaction, with cooperation as its principal motto. Here, the premise that companies are relational actors, as advocated by the theory of VoC, was taken to the letter by the group responsible for creating the model presented. Moreover, one can also say that FIRM's initiative corresponds positively to the proposal of the "New Legal Framework for Science, Technology and Innovation" in Brazil, which aims to reduce the gap between business and academic community.

However, as expected, the collaborative model still has much resistance, both internally and externally. Within the company, it is feared by the strategic thematic sharing. Competitiveness is still seen through the secrecy which, in practice, reduces the possibilities if we consider that the strength of this proposal is the formation of cooperation networks. Intellectual property agreements, commonly used by companies when they interact with universities, denote this point well. This aspect is an issue and should be consider in the next legislative discussions.

Another obstacle to the implementation of the collaborative model are ST&I assessment practices adopted by government agencies, which were not the subject of discussion in Law nº. 13:243. In Brazil, almost 100% of the scientific production takes place within universities and research centers. The public policies to increase production reinforce this aspect more "quantity" and less "cooperation – interaction". Thus, research coordinators are more encouraged to make publications and to form post graduate students, than to invest their time and availability to meet the modus operandi of industries and learn how to work with them. Here, we need institutional mechanisms to minimize the distance between these two worlds: the Science and the Innovation.

The university-company relationship in Brazil is marked by a history of disagreements. The approach between both is very recent, which contributes to the innovative enterprises remain almost unreachable by Brazilian society. A R&D management and evaluation model more collaborative and based on the complementarity of interests can contribute to that capitalist economies find new ways of survival to face institutional changes. The "New Legal Framework for Science, Technology and Innovation" (Law nº. 13.243) is a step in this direction. However, besides legislative changes, both companies and academic institutions should also be willing to change their behavior in order to create mechanisms which implement the approach between each other, as the example of the FIRM case presented here.

7. REFERENCES

Hall, P. and D. Soskice (2001): Varieties of capitalism. The institutional foundations of comparative advantage, New York: Oxford University Press.

Lazonick, W. (2007): "Varieties of Capitalism and Innovative Enterprise", in Lars Mjøset, Tommy H. Clausen (ed.) Capitalisms Compared (Comparative Social Research, Volume 24), Emerald Group Publishing Limited, p. 21 – 69.

Nazareno, C. (2016). "As mudanças promovidas pela lei nº 13.243, de 11 de janeiro de 2016 (Novo Marco Legal de Ciência, Tecnologia e Inovação) e seus impactos no setor". Estudo Técnico. Câmara dos Deputados, Brasília, DF.

North, D. (1998): "Organizations, learning and institutional change". In: Institutions, institutional change and economic performance, New York: Cambridge University Press.

Soskice, D. (1999): "Divergent production regimes: coordinated and uncoordinated market economies in the 1980 and 1990s". In: Kitshelt, H. et al. Continuity and change in contemporary capitalism, New York: Cambridge University Press.

World Bank. Research and development expenditure (% of GDP) | Data. Available in: <http://data.worldbank.org/indicator/GB.XPD.RSDV.GD.ZS?locations=BR>. Acess: October, 31th, 2016.

8. REFERENCES NOT CITED

Granoveter, M. (2005): "The impact of social structure on economic outcomes". In: Journal of Economic Perspectives, vol. 19, nº.1.

_______________ (1992): "Economic institutions as social constructions: a framework for analysis" In: Acta Sociologica. New York: State Univ. of New York, nº. 35, p. 3-11.

Schneider, B. R. (2004): "Varieties of semi-articulated capitalism in Latin America". In: Annual Meeting of the American Political Science Association, Chicago.

AUTHOR

DOMENICA BLUNDI
Avenida das Américas, 700/3º piso/Loja 318 – Barra da Tijuca
Rio de Janeiro – RJ – Brasil – cep: 22640-100

RESEARCH IN SOCIAL SCIENCES AND HUMANITIES (SSH) AT AUSTRIAN UNIVERSITIES: BIBLIOMETRIC ARTICLE ANALYSIS AND COMPARISON OF THE YEARS 2007, 2010 AND 2013

GÜNTHER R. BURKERT, DAVID F. J. CAMPBELL AND THORSTEN D. BARTH

How can research and research quality be measured? One dimension of manifestation of research quality focuses on the publication of articles in international journals that are being peer reviewed. Bibliometric analysis opens here a route of measurement of article publications. Based on such results, the formulation of propositions is possible, which can be connected to statements of evaluation.

In context of the here presented paper the results of a bibliometric study of article publications at Austrian universities in the social sciences and humanities (SSH) are being presented and will be further discussed with regard to their possible relevance for evaluation and governance (internal governance, external governance). The bibliometric study was based on the following methodic design: (1) Only articles were considered, with at least one institutional address in Austria and that were released in journals that are being represented either in the SSCI (Social Sciences Citation Index) or the A&HCI (Arts and Humanities Citation Index). This reflects the disciplinary spectrum of the social sciences and humanities (SSH). (2) Assignment and aggregation of articles to institutions was based and processed through the "institutional addresses" of the articles. (3) The following seven Austrian universities were considered for analysis: Universität Graz (Karl-Franzens-Universität Graz); Universität Innsbruck; Universität Klagenfurt (Alpen-Adria-Universität); Universität Linz (Johannes-Kepler-Universität Linz); Universität Salzburg (Paris-Lodron-Universität Salzburg); Universität Wien; und Wirtschaftsuniversität Wien (WU Wien). (4) Three years were compared: 2007, 2010 and 2013. As first preliminary results the following propositions already can be set up:

1. **Growth in general:** The total result (during the period 2007-2013) shows a definite growth of the published articles in the social and human sciences (humanities) that is even stronger in the social sciences (articles and citations). The human sciences (mainly in the citations) are ranked behind the social sciences.
2. **Growth in detail:** This is most evident in psychology and also on a few (but not in all universities) of the economic sciences (but not for the University of Vienna.) The departments of the human sciences have also expanded considerably, but by tendency more in the areas of article publications and less for citations. For the social and human sciences the University of Vienna has the strongest position and there was also the most obvious growth in Austria (for articles and citations). On the other hand, the University of Vienna does not demonstrate any growth in the economic sciences.
3. **No growth:** The jurisprudence departments show almost no growth in the discussed areas.

In the following, the analysis outcome for the University of Vienna is being summarized in greater detail:

1. The departments with strongest growth are: psychology (articles and citations), and the social sciences (articles and citations). Political science and journalism ("Publizistik") are the strongest in the social sciences. The significant growth of the departments for social sciences is obvious in the internal Austrian comparison across different Austrian universities.
2. The department of human sciences (historic-cultural sciences, philological-cultural sciences) also has an impressive growth, in which the citations are weaker in philology.
3. The department of economic sciences is partly stagnating.

The further development of the paper should address the following issues: (1) **a comparative commenting on the publication performance in the SSH for all seven Austria universities** that were covered by the analysis; (2) **Discussion of methodic design considerations** for bibliometric analyses; (3) **Discussion of possible propositions for (internal, external) governance at Austrian universities**. More specifically, the governance theme will address the following issues:

1. The particular situation of the social sciences and humanities (at Austrian universities);
2. Legitimacy aspects of the social sciences and humanities;
3. Consequences of increased competition;
4. The growing role of the importance of project-based funding ("Drittmittelfinanzierung");
5. "Quantitative" measures and references (goals) for performance;
6. Outcome and results of the governance of the state (responsible ministry) for the social sciences and humanities at the Austrian universities.

AUTHORS

DOMENICA BLUNDI
Avenida das Américas, 700/3º piso/Loja 318 – Barra da Tijuca
Rio de Janeiro – RJ – Brasil – cep: 22640-100

GÜNTHER R. BURKERT
Federal Ministry of Science, Research and Economy (BMWFW), Minoritenplatz 5, A-1014 Vienna
Austria
Guenther.Burkert@bmwfw.gv.at

DAVID F. J. CAMPBELL
Corresponding author
Alpen-Adria-University Klagenfurt, Faculty for Interdisciplinary Studies (iff), Institute of Science Communication and Higher Education Research (WIHO)
Schottenfeldgasse 29, 1070 Vienna
Austria
david.campbell@aau.at
Telephone: +43 699 10450856

THORSTEN D. BARTH
Federal Ministry of Science, Research and Economy (BMWFW)
Minoritenplatz 5, 1014 Vienna,
Austria
Thorsten.Barth@bmwfw.gv.at

FUNDING FRUGAL INNOVATIONS: AVAILABILITY AND DESIGN OF PUBLIC FUNDING PROGRAMMES FOR FRUGAL INNOVATIONS

KAISA GRANQVIST

A heart beats unmistakably on the monitor as Dr Devi Shetty, explains the results of a heart scan to a concerned patient. This patient is one of the many patients, treated in Dr Shetty's Narayana Hrudayalaya Group's hospitals spanning to 32 cities, towns and villages in India. Dr Shetty has been called Henry Ford of surgery by the Wall Street Journal (Anand 2009) as his hospitals provide world class cardiac care at radical low cost by applying the philosophies of mass production and lean manufacturing. Heart surgery costs between $2000 and $5000, compared with $20,000 to $100,000 in the US, and his goal is to reduce the cost to $800. In addition to a low cost, the business model allows offering more than 60 free surgeries for poor patients while having a higher profit margin than an average American hospital (Bound and Thornton 2012; Raghunathan 2015).

Dr Shetty's story is a striking example of frugal innovation. Frugal innovations can be defined as cost efficient innovations with a social aim, often originating from or targeting developing or emerging economies (e.g. Bhatti 2012). Frugal innovations are distinctive in both their means and their ends (Bound and Thornton 2012). They are distinctive in their means because they respond to limitations in resources, whether financial, material or institutional, and turn these constraints into an advantage. They also need to be compatible with the unique circumstances of the poorest population groups (Nakata 2011:3, Van Beers et al 2014:5). The nature of the services, products and processes is also distinctive (Bound and Thornton 2012). Successful frugal innovations are not only lower in cost but surpass or maintain performance dimensions and can be made available at large scale (Van Beers et al 2014). Frugality spans from product design to new innovative business models, service delivery approaches. Moreover, frugal innovations have often explicit social aim or create considerable social impacts (Bhatti 2012, Bound and Thornton 2012).

Looking back to the example of Dr Shetty's cardiac hospital, it can be understood that frugal innovations have not captured the attention of companies and non-profits operating at the developing countries but also policy makers and other innovation stakeholders in western economies, who face the ever increasing pressure of ageing society. Also environmental constraints around climate and energy will increase demands for frugal models or production and consumption. As Rao (2013) has stated: "Inventions that initially seek to meet the requirements of the poor population in developing countries eventually can help to solve some of the developed nations' most pressing environmental and social problems".

There is also another reason, which has encouraged western policy makers and businesses to turn their looks at the emerging economies, and increased the interest in frugal innovations. While the growth in many western economies has stagnated, the economic growth of many emerging markets is growing at around 8 percentages annually. This creates a clear economic incentive to tap into these markets, where operating often requires adoption of frugal principles.

Following this, Bhatti (2012) argues that the most fruitful ground for frugal innovations emerges in the intersection of business-driven, social and institutional innovations. Drawing from the theoretical differentiation of these innovation types (e.g. Basu et al 2013; Howald et al 2015; Technopolis 2012), different kind of frugal innovations emerge in the three intersections with different aims and motivations (Bhatti 2012):

- The frugal innovations in the intersection of business-driven and social innovation aim to address the problems of the poor and simultaneously create profit for a (western) business by developing new (frugal) products to low-income markets in developed at developing economies (Bhatti 2012; Radjou & Prabhu 2015) With the new products and services to low-income markets, the innovators aim to achieve cost leadership as a competitive advantage by generating high profits through low cost and high scalability (e.g. Nakata 2012, Nakata & Weidner 2012). Because the new products and services improve the well-being of the poor they are simultaneously socially relevant (Prahalad 2012).
- Frugal innovations in the intersection of business-driven and inclusive innovations put first the needs of the citizens at the bottom of the pyramid in order to develop appropriate, adaptable, affordable, and accessible services and products to respond their needs. In order to do this, social responsibilities are included in business strategies or conventional views to empower the poorest population groups are inverted. (Prahalad 2012).
- In the intersection of institutional and social innovations occur non-profit or local activities by the civil society to address the needs of the low-income groups. These innovations solve personal needs of users with little or no aspiration to profit or scale from the innovation (Bhatti 2014). These kinds of frugal innovations root to social movements, institutional entrepreneurship, non-profit activities, and new collaboration arrangements between actors and have been largely addressed by organizations

from international development actors to local NGOs (e.g Mair & Marti 2009)

The notion that frugal innovations emerge in the intersection of business-driven, social and institutional innovation suggests that frugal innovation process is constitutively different than for "ordinary" R&D driven innovations or social innovations (Van Beers et al 2014).

Therefore, this study aims to shed light on the motivations behind and the design of public funding instruments particularly focused on supporting frugal innovations. Because frugal innovations have emerged in the academic discussion (Prahalad 2004) only in the last ten years, we also explore the availability of public funding instruments for frugal innovations, particularly in the western countries. Lastly, as the literature suggest that frugal thinking is required more and more in all spheres of innovation, the study discusses the lessons learned from frugal innovations support instruments and their implications to instruments supporting social and business driven innovations in western economies. More precisely, the study aims at answering the following research questions:

- Are there public funding instruments which particularly focus on stimulating frugal innovations?
- What are the objectives and motivations of frugal innovations funding instruments?
- Does the design of the instruments acknowledge the specificities of frugal innovation processes?
    - How the design of frugal innovation instruments differ from the "traditional" technological innovation support instruments or more recent attempts to support social innovations in developed economies?

The study aims at answering the first research question these by a desk research; online search of public frugal innovation instruments. The second and the third research questions will be answered by conducting semi-structured telephone interviews with the owners of these instruments. The study defines a public support instrument as an instrument funded by a government or its intermediate (including international non-profit organisations with contributions from government public sources), however, excluding regional and local support instruments, although being funded and operated by a government organisation.

The paper seeks to make a theoretical contribution on frugal innovation theory, and particularly what is known of frugal innovation process and the availability of public funding instruments for frugal innovation.

REFERENCES

Anand, G. (2009). The Henry Ford of Heart Surgery. In India, a Factory Model for Hospitals Is Cutting Costs and Yielding Profits. The Wall Street Journal. Accessed 08 May 2016 < http://www.wsj.com/articles/SB125875892887958111>

Basu, R., P. Banerjee & E. Sweeny (2013). Frugal Innovation. Core Competencies to address global sustainability. Journal of Management and Global Sustainability 2 (2013) 63-82.

Bhatti, Y. (2012).What is Frugal, What is Innovation? Towards a Theory of Frugal Innovation. Academy of Management Annual Meeting, Boston 3-7 August 2012.

Bhatti, Y. (2014). Frugal Innovation. Social Entrepreneur's perceptions of innovations under institutional voids, resource scarcity and affordability voids. Unpublished dissertation. University of Oxford, the United Kingdom.

Bound, K. & I .Thornton (2012). Our Frugal Future: Lessons from India's Innovation System. 77p. Nesta, United Kingdom.

Howald J, R Kopp, & M Schwarz (2015) Social Innovations as Drivers of Social Change – Exploring Tarde's Contribution to Social Innovation Theory Building. In A. Nicholls et al. (eds.), New Frontiers in Social Innovation Research. 29-51pp.

Mair, J. & I. Martí (2006). Social entrepreneurship research: A source of explanation, prediction, and delight. Journal of World Business, 41(1): 36-44.

Nakata, C. & K. Weidner (2012). Enhancing New Product Adoption at the Bottom of the Pyramid: A contextualized model. Journal of Product Innovation Management, 29 (1), 21-32.

Prahaland, C.K. (2012). Bottom of the Pyramid as a Source of Breakthrough Innovations. Journal of Product Innovation Management, 29 (1), 6-12.

Radjou, N. & J. Prabhu (2015). Frugal Innovation- How to do more with less. Economist Books. 226pp. Economist, London.

Raghunathan, A. (2015). India's Affordable Healthcare Doyen – Dr. Devi Shetty Files for IPO. Forbes. Accessed 08 May 2016 <http://www.forbes.com/sites/anuraghunathan/2015/10/29/whats-in-the-ipo-filing-by-indias-affordable-healthcare-doyen-dr-devi-shetty/#2575748b181b>

Rao, B. (2013). How disruptive is frugal. Technology in Society 35 (2013) 65-73.

Technopolis (2011). Trends and Challenges in Public Sector Innovation in Europe. Thematic Report 2012 under Specific Contract for the Integration of INNO Policy Trendchart with ERAWATCH (2011-2012). Commissioned by the European Commission.

Van Beers, C., P. Knorringa & A. Leliveld (2014). Frugal Innovation in Africa: Towards a Research Agenda. Position paper 14th EADI General Conference | 14thEADI General Conference Responsible Development in a Polycentric World: Inequality, Citizenship and the Middle Classes

AUTHOR

KAISA GRANQVIST
Centre for Social Innovation, Austria
Linke Wienzeile 246, 1150 Vienna
Austria
Tel. 00436604743059

SAME OBJECTIVES, DIFFERENT GOVERNANCE – HOW THE EXCELLENCE INITIATIVE AND THE PACT FOR RESEARCH AND INNOVATION AFFECT THE GERMAN SCIENCE SYSTEM

KAISA GRANQVIST

INTRODUCTION

Research funding can be characterized as an instrument used by funding councils and science policy makers to affect the research of individual researchers, organizations or the whole research system. Research topics and funding schemes should be carefully chosen to achieve the funders' goals. More than ten years ago the German federal government and the states had the overall objective to strengthen the German science and university system and their international competitiveness by focusing mainly on research excellence. They initiated two large research funding programs: the Excellence Initiative (ExV) and the Pact for Research and Innovation (PFI). The two funding programs have both similarities and differences. While the Excellence Initiative is dedicated to the university system, the Pact for Research and Innovation focuses on the public non-university research organizations. Although the Excellence Initiative and the Pact for Research and Innovation pursue the same goal, different funding and governance mechanisms are applied. This leads to two questions: What are the reasons for choosing different forms of funding in order to fulfill objectives that are to a great extent identical? How do these differing governance mechanisms affect the universities and the public non-university science system? The effects will be observed by bibliometric and research and development (R&D) indicators.

FUNDING SCHEMES AND GOVERNANCE MECHANISMS

The governance perspective has increased continuously over the last decade and has influenced the research on higher education and science studies. For instance, the model of the "Governance Equalizer" (Boer, Enders, & Schimank, 2007; Schimank, 2005) differentiates analytically between five dimensions of governance: state regulation, stakeholder guidance, academic self-governance, managerial self-governance and competition. In an international comparison the German higher education system could traditionally be characterized by both a relatively strong state regulation and a relatively strong academic self-governance, but an undergoing profound process in the last years has transformed the universities more and more into organizational actors (Krücken, 2011).

What is the underlying funding and governance mechanism of the Excellence Initiative and the Pact for Research and Innovation? The governance of the Excellence Initiative is based on competition. The science policy aim of the program was to promote a "performance spiral" (ExV), which should lead to a higher performance und a better international standing of the German universities. Proposals for competitive grants have to be submitted and are reviewed in a group peer review process. The highly selective funding scheme (Möller, Antony, Hinze, & Hornbostel, 2012) produces temporarily funded winner universities. In contrast, the central science policy aim of the Pact for Research and Innovation was to give the public research organizations financial planning security, which means that the block grant steadily rises for the public non-university research organizations by an annual rate of 3% (2006-2010 and 2016-2020) respectively 5% (2011-2015).

It begs the question why the federal government and the states decided to run different funding schemes (competitive versus block grants) in the two programs which have largely the same goals, to strengthen the German science and university system and their international competitiveness.[1] Two developments framed the conceptualization phase of the Excellence Initiative: An ongoing debate since the 90s about the "rotten" German universities (Glotz, 1996; Simon, 1991) and the results of the international university rankings (Shanghai 2003, THE 2004), which show that the German universities couldn't keep up with the top 50 worldwide leading universities. During this time the managerial self-

1 In addition, the Excellence Initiative and the Pact for Research and Innovation supports collaborations between the university's and the non- university research organizations as well as promoting young researchers and gender equality. The Pact for Research and Innovation gives a somewhat greater weight on the cooperation with the industry.

governance of universities was highly questionable. It seemed that the competitive governance mechanism had to be the modus operandi for giving new impulses to the German university system – according to the ideas of New Public Management. In contrast, it looked as though the federal government and the states had a deeper trust in the managerial self-governance of the non-university research organizations, which also have a higher research performance than the universities (see below). The governance mechanism of the Pact for Research and Innovation can be described as an external state guidance in terms of a target agreement, in which the science policy sets the main objectives, but give the non-university research organizations the freedom to choose (GWK, 2014) the appropriate activity for achieving the given goals. Every year the non-university research organizations have to report their annual activities, but the future amount are independent of the degree of target achievement.

METHODOLOGY

The paper is based on data of the German Federal Statistical Office (Statistisches Bundesamt) and bibliometric data of the Web of Science (WoS). The development of the research and development expenditures and the third-party funding were analyzed. In addition, data from the monitoring report of the Pact for Research and Innovation (GWK, 2014, pp. 77–79) is included, because the Federal Statistical Office didn't annual report the third-party funds of the non-university research organizations. For the bibliometric analysis the publications of the humanities and social sciences were excluded, due to a very low coverage in the WoS. For the analysis of the proportion of publications that belong to the worldwide top 10% highly cited papers (PP Top 10%, Bornmann, 2014; Waltman & Schreiber, 2013) only citable document types (articles, reviews) were considered and measurement field and doctype normalized in a three year citation window. All calculations have been processed on the database of the German Competence Center for Bibliometrics.

The study covers the time period from 2005 to 2012. The year 2005 was the last year before the financial support from both the Excellence Initiative and the Pact for Research and Innovation started. The investigation period ends 2012 because the first funding period of the Excellence Initiative terminated in this year.

In order to differentiate the effects of the two funding programs different units were analyzed separately: One the one hand the universities and on the other hand the non-university research organizations (Helmholtz Association (HGF), Max Planck Society (MPG), Leibniz Association (WGL) and Fraunhofer Society (FHG). In addition three university subgroups were separated by their success in the Excellence Initiative. First, the nine so called Universities of Excellence (UoE), which are successful in each of the three funding lines (the graduate schools, the cluster of excellence and the university future concept), second, the 37 universities with success in at least one funding line (ExIn Uni) and third, the universities whiteout any funding in the Excellence Initiative (No ExIn Uni). For the bibliometric approach only eight Universities of Excellence (UoE) were considered, because a separate analysis was not applicable after the merger between the University of Karlsruhe and the Helmholtz Center Karlsruhe. The address normalization were made in the German Competence Center for Bibliometrics (Schwechheimer, Rimmert, & Winterhager, 2015).

RESULTS

Research and development expenditures and third-party funding

How do different governance mechanisms affect the universities and the non-university research organizations? Figure 1 shows the development of the R&D expenditure and Figure 2 the third-party funding of the universities (without the university hospitals) and non-university research organizations (NURO). The data were normalized to the year 2005.

Figure 1 show similarities and differences between the university and the non-university sector. During the time period from 2005 to 2012 the growth of R&D expenditure for university and the non-university sector are almost the same (Uni 150%; NURO 148%). While the funding by the Pact for Research and Innovation fully comes to bear in 2006 the financial support of the Excellence Initiative rises slowly over the first two years. Some Initial projects started in November 2006 and the last approved projects began their work at the end of 2007. So 2008 is the first year in which all the projects of the Excellence Initiative were running.

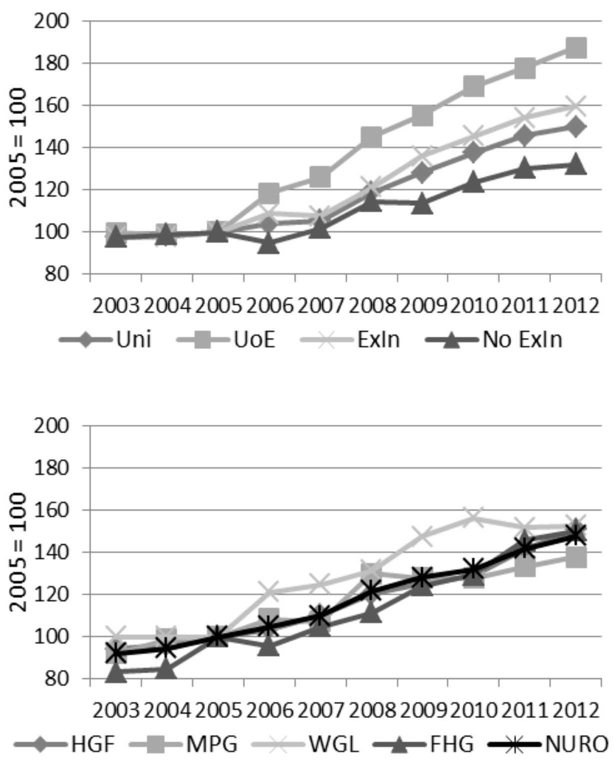


Figure 1: Development of the R&D expenditures of the university groups (w/o hospitals) and the non-university research organizations (NURO)

The increase of the universities of excellence (UoE 187%) and the excellence universities (ExIn 160%) are above the average while the non-excellence universities have a lower growth rate (No ExIn 132%). These results indicate an ongoing stratification process in the German university system. In comparison the increase of R&D expenditures of the non-university research organizations are more similar: WGL 153%, HGF 151% and the FHG 150%. Only the MPG has a slightly lower growth rate (138%).

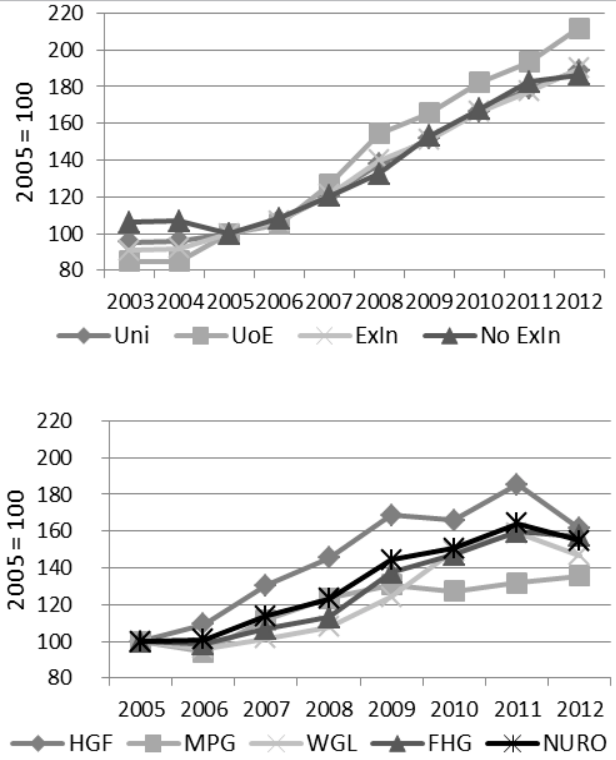


Figure 2: Development of the third-party funding of the university groups (w/o hospitals) and the non-university research organizations (NURO)

The third-party funding (Figure 2) of the university groups rise more sharply than the non-university research organizations (Uni 189%; NURO 155%). The UoE (212%) have the highest growth rate in the university sector, but the overall stratification in the university sector is not as distinctive as the R&D expenditures (Figure 1). The universities with and without an excellence funding have an almost similar increase (ExIn 190%, No ExIn 186%). The third-party funding for non-university research organization – except the MPG – decreased sharply in 2012, because a federal funding program supporting research infrastructures in the non-university research sector during the financial crisis (2009-2011) was terminated. The HGF has the highest growth of third party funding (161%) followed by the FHG (157%), the WGL (147%) and the MPG (136%).

BIBLIOMETRICS

In the bibliometric analysis (Figure 3) the indicator proportion of publications that belong to the top 10% worldwide highly cited papers (PP Top 10%) is applied. The university groups show the above mentioned differentiation: The UoE (year 2011: 17.1%) is exceeding the ExIn universities (15.5%) followed by No ExIn Universities (13.2%). Above the overall German average (14.8%) are the UoE, and the ExIn universities.

The annual results for the FHG differ over time by a low number of publications, but it should be noted that publication or citation indicators are not very appropriate for an organization, which has a strong applied research and industry collaboration mission. The best performers in the German research systems in 2011 measured by the PP Top 10% indicator are the MPG (23.0%), HGF (19.8%), UoE (17.1) and the WGL (16.9%). The non-university research organizations have overall higher impact (NURO 19.6%) than the university sector (14.8%).

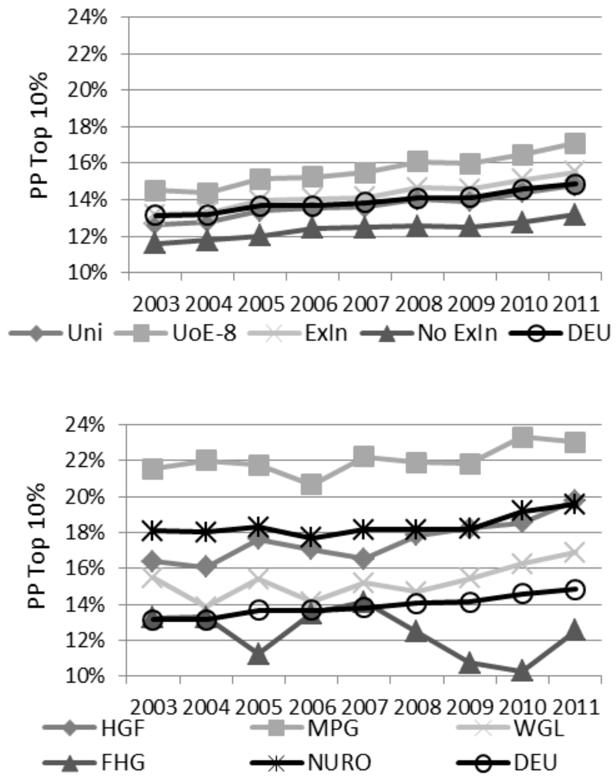


Figure 3: Proportion of publications that belong to the top 10% worldwide highly cited papers (PP Top 10%) of the university groups (w/o hospitals) and the non-university research organizations (NURO)

DISCUSSION

The Excellence Initiative and other third party funding programs have changed the research conditions from the universities. The third-party funding rises more sharply than the total R&D expenditures (Figure 1 & 2). As shown in Table 1 in 2005 39% of the R&D expenditures of the university sector based on third party funding. In 2012 it rises up to 49%. The results indicate where a different governance of funding leads to: The competitive funding enhances the share of third party funding, while a growth in basic funding has the converse effect. For the non-university sector with a steady increase in basic funding the share of third party funding stays almost at the same level (2005: 31%; 2012: 33%).

	Uni	UoE	ExIn	No ExIn	NURO	HGF	WGL	MPG	FHG
2005	39%	51%	44%	31%	31%	23%	23%	17%	64%
2012	49%	58%	52%	44%	33%	25%	26%	16%	70%

Table 1: Share of third party funding of the total R&D expenditures of the university groups (w/o hospitals) and the non-university research organizations (NURO)

The university groups have a higher share of third party funding than the non-university research organizations, except the FHG. In comparison with the bibliometric results it can be concluded that a high share of third-party funding doesn't necessarily correlate with a higher value of PP Top 10% indicator. The MPG with the lowest share of third party funding (2012: 16%), is the outperformer of the German research system. For a valid comparison between the universities and the non-university research sector further aspects should be taken into account, e.g. discipline related financial demands, the industry mission orientated research (e.g. of the FHG), or large research infrastructure (e.g. German Electron Synchrotron (DESY) as a part of HGF). All these factors limited the direct comparison between universities and non-university research organizations.

Within the university groups the UoE have the highest share of third party funding (2012: 58%) followed by the ExIn (52%) and the No ExIn (44%) universities. Considering that the third party funding doesn't cover the whole research expenditures, the success in attracting additional funds may yields, especially for the most competitive and successful universities, into internal governance problems. A report of the German Research Foundation (DFG) stated that depending on difference between the research fields and topics from 30% up to 300% of the personal costs of a research project has to be co-financed from university (DFG, 2013). That's over, in part far over the given flat rate of 20% overhead budgets provided from the German Research Council.

Former findings based on guided interviews with university leaders and researchers in the Excellence Initiative show that with a growing amount of co-financing the universities are more and more restricted in their future opportunities of actions (Bukow & Möller, 2013). But, the ability to act is essential for the organizational self-governance of a university. As some studies pointed out, organizational autonomy is an important factor for success in attracting competitive funds (Aghion, Dewatripont, Hoxby, Mas-Colell, & Sapir, 2010; Boer, Jongbloed, Enders, & File, 2010). An extensive competitive funding that limited the self-governance of the universities via a high degree of not fully funded research can have in the end no or the opposite effects. In contrast to the universities the non-university research organizations have a greater ability to choose the appropriate activity also within the funding scheme Pact for Research and Innovation.

Currently, we are facing a new trend in some German states (e.g. Baden-Württemberg, Thüringen). They have reduced their competitive and program based funding in favor of an annual increase in basic funding. Similarly to the Pact for Research and Innovation the state Baden-Württemberg assure an annual increase of 3% until 2020. The explicit aim of the science policy makers is to give the universities more flexibility and autonomy and less state regulation (Baden-Württemberg, 2015, p. 6). After a period focusing mainly on the governance mechanism competition science policy makers seems to rethink their toolbox instruments and emphasis also other factors that are relevant for the university self-governance, their autonomy and productivity.

ACKNOWLEDGEMENTS

This study was funded by the German Expert Commission for Research and Innovation and the German Federal Ministry of Education and Research. All calculations have been processed on the database of the German Competence Center for Bibliometrics.

REFERENCES

Aghion, P., Dewatripont, M., Hoxby, C., Mas-Colell, A., & Sapir, A. (2010). The governance and performance of universities: evidence from Europe and the US. Economic Policy, 25(61), 7–59. https://doi.org/10.1111/j.1468-0327.2009.00238.x

Baden-Württemberg (2015). Perspektiven 2020. Hochschulfinanzierungsvertrag Baden-Württemberg 2015-2020. Retrieved from https://mwk.baden-wuerttemberg.de/fileadmin/redaktion/m-mwk/intern/dateien/publikationen/411503_MWK_Infografiken_Broschuere_RZ_onl.pdf

Boer, H. de, Enders, J., & Schimank, U. (2007). On the Way towards New Public Management? The Governance of University Systems in England, the Netherlands, Autria, and Germany. In D. Jansen (Ed.), New Forms of Governance in Research Organizations, 137–152. Dordrecht: Springer.

Boer, H. de, Jongbloed, B., Enders, J., & File, J. (2010). Progress in higher education reform across Europe, Funding reform, Volume 1: Executive Summary and main report. Enschede: Centre for Higher Education Policy Studies (CHEPS). Retrieved from http://doc.utwente.nl/88692/

Bornmann, L. (2014). How are excellent (highly cited) papers defined in bibliometrics? A quantitative analysis of the literature. Research Evaluation, 23(2), 166-173.

Bukow, S., & Möller, T. (2013). Die Rekrutierung wissenschaftlichen Spitzenpersonals in der Exzellenzinitiative. Berlin: Institut für Forschungsevaluation und Qualitätssicherung (IFQ).

DFG (2013). Bericht der Deutschen Forschungsgemeinschaft über die Erfahrungen mit der Gewährung der Programmpauschale in der zweiten Programmphase. Retrieved from http://www.gwk-bonn.de/fileadmin/Papers/DFG-Bericht-2013.pdf

ExV, Exzellenzvereinbarung (2005): Bund-Länder-Vereinbarung gemäß Artikel 91 b des Grundgesetzes (Forschungsförderung) über die Exzellenzinitiative des Bundes und der Länder zur Förderung von Wissenschaft und Forschung an deutschen Hochschulen. 18.07.2005. BAnz S. 13347.

GWK (2014). Pakt für Forschung und Innovation Monitoring-Bericht 2014. Bonn: Gemeinsame Wissenschaftskonferenz (GWK).

Glotz, P. (1996). Im Kern verrottet?: fünf vor zwölf an Deutschlands Universitäten. Stuttgart: Deutsche Verlags-Anstalt.

Krücken, G. (2011). A european perspective on new modes of university governance and actorhood. Research & Occasional Paper Series, 17(11), 1–11.

Möller, T., Antony, P., Hinze, S., & Hornbostel, S. (2012). Exzellenz begutachtet: Befragung der Gutachter in der Exzellenzinitiative. Berlin: Institut für Forschungsevaluation und Qualitätssicherung (IFQ).

PFI, Pakt für Forschung und Innovation (2005). Retrieved from: http://www.gwk-bonn.de/fileadmin/Papers/pakt_fuer_forschung_und_innovation.pdf

Schimank, U. (2005). A Comparative Perspective on Changes in University Governance in Europe. Public Lecture at The Australian National University on 17.October 2005

Schwechheimer, H., Rimmert, C., & Winterhager, M. (2015). Dokumentation zur Institutionenkodierung. Version 1.4. Bielefeld: Kompetenzzentrum Bibliometrie.

Simon, D. (1991). Die Universität ist verrottet. Der Spiegel, 1991(50/1991), 52–53.

Waltman, L., & Schreiber, M. (2013). On the calculation of percentile-based bibliometric indicators. Journal of the American Society for Information Science and Technology, 64(2), 372–379.

AUTHOR

DR. TORGER MÖLLER
German Centre for Higher Education Research and Science Studies (DZHW)
Schützenstr. 6a, 10117 Berlin
Germany
e-mail: moeller@dzhw.eu

EVALUATING THIRD MISSION ACTIVITIES – TOWARDS A CONCEPT FOR SMALL UNIVERSITIES OF APPLIED SCIENCES

GABRIELE PERMOSER AND HANNES RAFFASEDER

The paper refers to the strong need of more diverse sets of key performance indicators for evaluating higher education institutions (HEI) due to the HEI landscapes' recent developments with its broaden variety of different institutional types and increasingly diverse strategic profiles. Especially, it discusses the main challenges of evaluating higher education institutions' third mission activities and points out the specific challenges of smaller universities of applied sciences. It outlines a holistic approach of integrating the three missions, higher education, research and knowledge transfer, suggest a scheme of convenient performance indicators and gives an overview about St. Pölten University of Applied Sciences concept for evaluating third mission activities.

Evaluating university research and teaching has already introduced and produced a variety of key data. Number of publications, impact factors, third-party funding, awards and patents are only a selection of relevant performance indicators, which are used to evaluate and rank university research (Campbell, 2005). The number of students, alumni, study programmes and teaching staff on the other hand are only a view of possible facts that are evaluated and compared in the field of teaching. These data are not only used to evaluate the institution itself but also the individual scientist (e.g. H-factor). Until now, all higher education institutions are evaluated with about the same key data and the same or at least very similar scales are used. However, during the last few decades the landscape of higher education has developed especially in view of quantity, specific profile of the different institutions and thus in diversity. While for instance Austria had 18 universities until 1990, there are now 21 universities, 12 private universities, 1 university for continuing education and 21 universities of applied sciences. A lively debate about strength and weakness, opportunities and threats of the different types and profiles of higher education institutions is going on accompanied by strong discussions about efficiency and funding schemes. According to the Austrian higher education institution plan for example different types of institutions should of course focus on their strengths, differentiate their profiles and cooperate within strategic partnerships (see Hochschulkonferenz). It is evident that these overall targets of differentiation and specialization would need a diverse set of key performance indicators, too. But up to now more or less the same type of data is used to measure the performance of different types of higher education institutions.

Especially the implementation of universities of applied sciences has not only added new actors to the system, but – as focused on practice-oriented education based on up to date scientific knowledge and on applied research and development instead of fundamental science – they have other targets, duties and responsibilities as well. Of course universities of applied sciences have installed their own research institutes and their continuously increasing research activities are in many cases are strongly related to industry, economy and society due to specific focus and the legal mandate of this type of higher education institutions. As the objectives of applied research are often close to the needs of the market, it is not in every case intended and sometimes even impossible to publish the results in peer-reviewed journals or to present papers at scientific conferences.

However, does it then make sense to evaluate the higher education institution performance with the same data if they are fundamentally different? Furthermore, the definition of certain key data motivates higher education institutions to concentrate on the improvement of their performance only based on this key data. Universities are using ranking results for attracting new students and excellent researchers. On the other hand, rankings can draw a picture of university performances based only on certain, not explained data, which are used from the public to criticize them.

Higher education institutions have been facing an ongoing discussion about their changing role in society and economy. After Wilhelm von Humboldt's idea of a university based on principles of education through learning and research, unity of research and teaching and academic freedom of research, teaching and learning dominated the evolution of higher education especially in Europe for about 200 years, the concept of the so-called "third mission" came up a little longer as a decade ago. It became rapidly popular within the academic discourse to point out that higher education institutions have to add a further goal in addition to the two missions of "teaching" and "research". While governments adopted their policies on higher education, academics have been hotly debating the integration of this "third mission" for several years now. However, in too many cases it is still a rather flexible and vague concept that can support entrepreneurial and commercial activities and serve as an additional funding stream as well as promoting a university's effort to take more societal responsibility. Thus, many higher education institutions still discuss their positions and struggle with the practical implementation of theoretical approaches and strategies, while new challenges, trends, technological changes and (maybe) buzzwords are coming up almost every month. (see also Benneworth, et al, 2009; 3M Project, 2012). Of course policy-makers and experts have suggested indicator systems, but due to the lack of an explicit description of "third mission" and the wide variety of its activities a constructive and objective evaluation is still a difficult task.

The St. Pölten University of Applied Sciences with it's about 2600 students is a rather small institution compared to other higher education institutions on international level. Located in the rather small town St. Pölten, the main capital of the federal state of Austria, Lower Austria, St. Pölten University of Applied Sciences (St. Pölten UAS) is both regionally anchored and globally connected. Third mission activities have already been on its agenda for several years. These activities range from further education trainings to internationally recognized science communication events, e.g. European Researchers' Night to workshop and trainings for pupils (e.g. vifkids summer academy) and elderly people (e.g. Senioren-Uni) to specific programmes to strengthen the entrepreneurial skills for both employees and students (e.g. creative pre-incubator). Researcher at the St. Pölten UAS are asked to build individual interactive exhibits for nationally and internationally recognized museums (e.g. Festspielhaus Baden-Baden, Österreichische Nationalbibliothek, Landesmuseum Niederösterreich). In addition to these activities, students are trained in 17 study programmes in six different departments ranging from media and technology, media and economy, railway infrastructure, information security, health and social sciences. Six research institutes in the fields of research mentioned before are conducting state of the art applied research, often in direct cooperation with both international and regional companies, administration, social services, etc.

Considering interaction with these manifold stakeholders St. Pölten UAS aims at a holistic and integrative approach of its three missions of teaching, research and knowledge transfer and targets to develop towards a platform for collaborative innovation. This platform has to include and interact with

- a divers variety of institutions and organizations, such as global companies as well as regional SMEs and start-ups, NGOs, schools and training centres, science centres and museums, coworking spaces and innovation hubs, accelerators and incubators, science and business parks, research labs and other higher education institutions, regional and federal governments, interest and pressure groups, funding agencies and others
- divers variety of ambitious and smart individuals, such as pupils, students and alumni, lecturers, scientists and developers, artists, prac-tioniers and industry experts, business angels and investors, decision makers, responsible citizens
- a divers variety of various disciplines and industrial sectors
- a divers variety of business models and processes as well as learning, teaching, research and innovation methods especially emerging from digital technologies

This interaction with various stakeholders demands different communication and dissemination strategies and activities and these activities need to be adapted to the target group (see also Gervais et. al, 2016). The holistic approach as well as the interrelation of different activities require performance indicators, which are suitable to present these interdependences, instead of a separate measurement of research output and quality aspects of teaching. Thus St. Pölten University of Applied Sciences has developed a matrix of manifold key performance indicators considering not only the output within the three missions, but their interdependencies as well. This model does not only sum the third-party funding and the peer-reviewed papers, it also includes the integration from students in research projects, the knowledge transfer to different audiences, different forms of knowledge transfer (e.g. workshops, articles in local newspaper, etc.), interaction with society as well as sustainability both for teaching and further research projects. In addition to that, there are further key criteria for evaluating third mission, especially in the field of science communication. In our evaluation concept, we include a closer examination of the target groups. How many different target groups (from kids to elderly people) can we reach with our activities? Do the activities focus only on one specific target group or are they broader oriented? Are there any developments at the projects resulting from science communication and knowledge transfer? For example, did investors or business became aware of research results and projects for further realization? How are public science/citizen science, open innovation, responsible science considered in research and dissemination activities? How can citizen science be presented in the performance reports? How can different stakeholders participate? How can the institution itself raise the awareness of dissemination of research results? How can these activities not only be included in the performance reports of the individual institution but also at the quantifying the output of the individual researcher?

Furthermore, St. Pölten UAS takes a closer look on initiatives for fostering innovative thinking and entrepreneurial skills for both students and employees. For example, a specific training concept for scientific staff had been developed and is currently implemented. Students are able to apply at the creative pre-incubator with innovative ideas, often developed within lectures and seminars. How can these initiatives influence the institutions performance and in addition to that influence its third mission activities?

The paper will provide an overview and a classification of the different key performance indicators to evaluate a wide range of activities, services and formats, which integrate higher education, research and knowledge transfer, connect various stakeholders and support the needs of different target groups. It outlines specific aspects of the evaluation of a small university of applied sciences' third mission activities and presents the practical approach of St. Pölten University of Applied Sciences.

REFERENCES

Benneworth, P., D. Charles, C. Conway, and P. Younger (2009): Characterising modes of university engagement with wider society. Newcastle: Office of the Pro-Vice-Chancellor

Campbell, D.F.J. (2005): Evaluation of University Resarch in Europe. Conceptual Framework and Comparative Typology. HOFO Working Paer Series 05.002. Vienna: IFF (Faculty for Interdisciplinary Studies) [online] available from http://www.iff.ac.at/hofo/WP/IFF_hofo.05.002_campbell_research_evaluation.pdf [3 May 2016]

E3M-Project (2012) Needs and constraints analysis of the three dimensions of third mission activities [online] available from http://www.e3m-project.eu/docs/Threedim-third-mission-act.pdf [2 April 2016]

Gervais, M., C. Marion, C. Dagenais, F. Chioccio and N. Houlfort (2016): Dealing with the complexity of evalutating knowledge transfer strategies: Guding pricnples for delveloping valdi instruments. Research Evaluation , 25 (1), 2016, p 62-68.

Hochschulkonferenz (o.A.) Weiterenwicklung des österreichischen Hochschulraumes [online] available from http://www.hochschulplan.at [3 May 2016]

AUTHORS

GABRIELE PERMOSER
St. Pölten University of Applied Sciences
Matthias Corvinus-Straße 15
3100 St. Pölten
Austria

HANNES RAFFASEDER
Corresponding author
St. Pölten University of Applied Sciences
Matthias Corvinus-Straße 15, 3100 St. Pölten
Austria
+43 2742 313 228 616

π-TUPI: AN OPENSOURCE P2P SOLUTION TO FOSTER OPEN EVALUATION?

DIEGO GOMES TOSTES, LUIZA ROSÂNGELA DA SILVA, JOSUÉ LAGUARDIA, JULIANA GONÇALVES DOS REIS, ELIAS RODRIGUES DE OLIVEIRA FILHO, ROBERTA LOUREIRO BARDUSCO AND MARCELLY MACHADO

INTRODUCTION

This article´s extended abstract is dedicated to present π-Tupi[1], an open source P2P platform for exchanging messaging, articles and comments between researchers and others interested in RTI. First of all and in essence, it is meant to be a call for collaboration and ample reflection for the conception and development on this Brazilian initiative still undergoing studies but which experimental initial module is already operational and accessible. The text carries, within the typical limitations of a project still in its early stages, comparisons with other initiatives that, much like π-Tupi itself, may be understood as a contribution to the discussions and efforts to advance in fostering the dimensions of transparency, collaboration and evaluation associated (or desirable) to the concept of Science 2.0. It is accepted as a principle that all these dimensions are contained in what, in general, is called open access to scholarly/scientific/technological/innovative communication, whether the intention is to present it, consume it or evaluate it.

The text also presents some of the specific characteristics of the π-Tupi platform that can help promoting open evaluation. Open and free access to scientific and technological communication and information is stated as universally positive goal. Nevertheless, it is also a slow and complex process, with numerous technological, economic and social setbacks and complications. In a recent post published by Pandelis Perakakis (1), one can find an illustration (Figure 1) that very simply and clearly sums up what in reality is expressed historically through disputes and negotiations among multiple actors, organized in multiple dimensions that rearrange themselves as technology advances and allows new ways to make RTI scholarly communication public.

The figure allows for quickly identifying that the dimensions described in the bottom line of the chart (dissemination/publication/limited accessibility) has seen effective progress in the last twenty years, primarily as a result of innovations in information and communication digital technologies, ie. they are not specific to the Academy. Precisely for this reason, the other two have not advanced in the same pace, hence the situation remains practically unaffected in what is RTI-métier specific. Not even the issues that are changing within this very métier are globally reflected in the processes and issues described in the illustration. For example, the fact that the RTI communication and research itself are

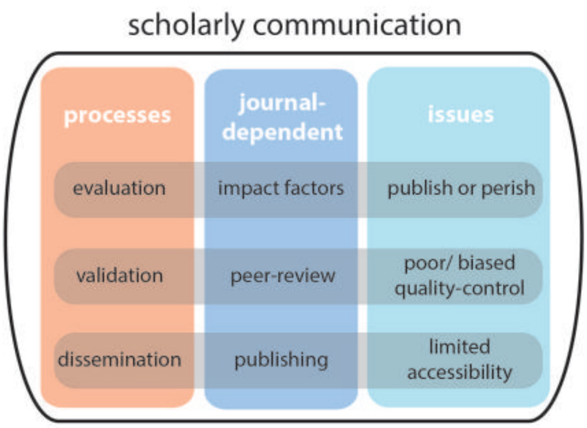


FIGURE 1
Essential scholarly communication processes controlled exclusively by journals
Source: Perakakis, 2016.

1 Pronounced is Portuguese as "peytopey". Pi, the number, is put together with Tupi, a Brazilian indian tribe in pre-colonial Brazil, so it can sound as "peer to peer". Tupi started an on-foot march from Amazonia to Brazil´s south leaving small portions of their own people behind to keep each portion of land they reached and, by doing so, they have spread all over the country´s inner and coastal lands. They became the biggest indian nation when Brazil was discovered by Portuguese navigators.

increasingly becoming a collaborative endeavour with no geographic nor disciplinary boundaries has not yet inspired the change in the organization of publishing RTI findings; it has paradoxically added new ethical and technical issues to the overwhelming need for a broad revision of that system.

There is, however, a number of theoretical proposals and approaches (2)(3)(4), social network platforms (5) (6) already in place or unfolding to make improvements in the alternative evaluation and validation of RTI outputs; those have gained either a lot of prestige or controversy (7) – in any case an enormous adherence of the academic community, but not as much from funding agencies and big publishers. In Brazil, there are many initiatives undertaken by Ibict (8) (9) in terms of open access publishing, but there are still few targeting the other two dimensions. A recent initiative aims to thrive in such direction: π-Tupi platform (pronounced peytopey in Portuguese).

In the words of its creator, systems analist Diego Tostes, the platform was born with the goal of providing a way to exchange articles and research „as good and popular as The Pirate Bay." Answers to this seemingly simple definition from a technical point of view - is it not just a P2P system, anyway? - guard, however, possibilities more complex as the more you expect the platform to get, in a context that makes it inevitable to compare π-Tupi with systems similar in goals, possibilities or limitations. The development/design team is part of the staff responsible for the Brazilian Clinical Trials Registry- ReBEC, the only open source clinical trials primary register platform in the world that contemplates all World Health Organization´s standards.

The π-Tupi platform perspective is based on the decentralization of data collections processing and availability. It consists of a simple web system with a database that stores only the metadata of any RTI output file (as an article, an interview, raw data from a survey and so on) and a torrent file generated by the author of the article, with which users can download data packages with outputs of RTI in different formats, not restricted scientific articles. P2P technology is used to decentralize the storage and to generate reports on the popularity of items: files with more „seeds" are more relevant according to the researchers themselves, which can make their own rating. The solution allows for code audit, while users can maintain confidentiality about their personal information, or not.

The metrics that π-Tupi can generate are still under study and development, but for now it can be said that they differ from those currently associated with the modalities of green, platinum and gold OA. Indexing allows for the generation of qualitative and quantitative metrics because the platform can be organized and seen as a social network - categorizable for information posting and retrieval, according to such culture, aesthtetics and technology, that is always platform-dependent. This creates some issues regarding the continued information. π - Tupi, differently, creates data exchanging conditions regardless of its platform , because its operation is distributed in terms of processing and persistence (9).

Its structure for information organization and exchange can be envisaged to generate reports in order to assist in understanding the social impact of prestigious connections between peers and different reputation building mechanisms. It should address the disciplinary communities where a particular publication or post originates from, as well as its reception in other fields of knowledge. Another subproject should regard opening access and categorizing the membership or temporary use of other agents vital to the RTI scenario, as policy makers, advocacy groups and social movements, industry or the individually concerned citizens.

MAIN OBJECTIVE

Opening the critical and technological development of the Open Source P2P platform called π-Tupi to the international academic community's inputs and scrutiny, aiming to contribute to the discussions and advances in the dimensions of transparency, collaboration and evaluation associated or desirable to the concept of Science 2.0.

SPECIFIC OBJECTIVES

- Make a survey of current demands in terms of advances in generating metrics for RTI evaluation.
- Make a survey on that platform in this context can be considered similar, checking contact points, differences and lessons learned.
- Characterize the possibilities and limitations unique to the π - Tupi platform.

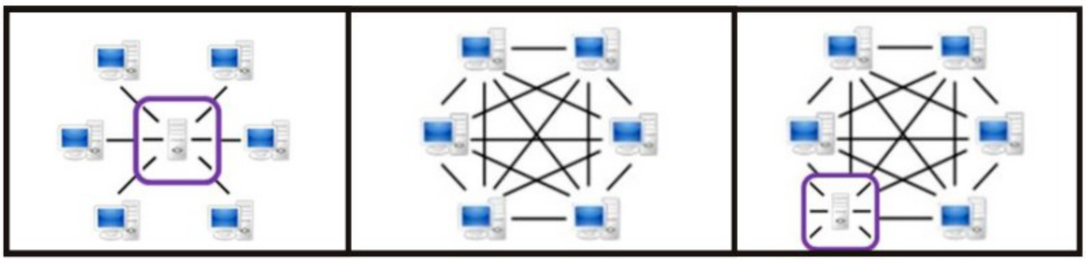


FIGURE 2
Concentrated servers versus Distributed operation on personal computers.
Distribution allows concentrated solutions, as repositories and social networks to become users of the same solution without demanding heavy investments in infrastructure

Source: Tostes et al, 2015.

METHODS

The methodology consists of reading recent texts that discuss the current situation and the future of transparency, collaboration and evaluation in RTI; a small survey on technological development and information management ongoing initiatives that may be considered similar to π-Tupi according to one or more of those three dimensions, as they contribute to the consolidation of oppenness as a principle and as a goal.

DISCUSSION

In this section, the article will address the history and the advantages of the open source P2P solution in the context of expanding access to scientific and technological information and communication. It then contextualizes the design and development of π-Tupi, its potential for generating metrics alternative or complementary to current evaluation models and platforms for researchers to present their work interact, cooperate, build and measure reputation – thus managing to strengthen the disciplinary and multidisciplinary scientific development but also their careers, individually.

A critical aspect of this solution , which will undoubtedly require specific further studies, concerns the model does not proposing the open evaluation by society and peers as a contradiction nor substitution to the traditional „blind" peer review, as it can preserve users' information anonymous or not. In fact, π-Tupi will make possible to evaluate whether or not identity is declared. The central efforts therefore can lie not in opening reviewers' identity data or not, but making reviewing itself a broader process which also protects other platforms (magazines, social media, repositories and other databases) from fraud, assessment biases or rework.

Such flexibility, given also by open source, means the review process could be a customized one, depending on the demands posed to the platform. A diversified review process does not need even to be understod only in terms of different knowledge approaches, as to take in consideration that RTI is each day becoming more collaborative. It means also providing a socially enriched framework whenever needed, in a way that social control of RTI and societal impacts can be facilitated to support funding and policy making decisions.

RESULTS

A preliminary list of initiatives was made using those three dimensions of OA as focal points, from transparency to technology issues. It will be necessary to refine and expand both the conceptualization and the quantity of criteria used to make this selection – a delicate task in which other researchers' would be most appreciated as corresponding to the critique of π-Tupi ongoing design. A summary panel was formed (Annex 1, Figure 3) taking into consideration the criteria so far pinpointed to help prognosticating the eventual role for π-Tupi in the current context of transparency, collaboration and evaluation in RTI.

Having this panel as a background, it is possible to explore some topics seeking to map out π-Tupi unique appeals and capacities, foremost of all there is the fact it is an open source P2P solution. It allows for distributed processing of information at various points of retrieval, and it is fully consistent with an open, decentralized collaboration model a among researchers. It also favours its improvement in terms technological development.

The model does not restrict depositing, collecting and assessment of information to a specific country or region or to any disciplinary area, though it allows to do so; therefore it should be possibel to draw rankings or peer evaluation taskforces per geographic area or knowledge. Contrary to the limitations imposed by current quantitative models, one could create relevant, fully customized instant reports also on transdisciplinary studies' impact.

FINAL CONSIDERATIONS

The initial, exploratory tests and projections made by the authors regarding the possibilities of π-Tupi indicate its potential to contribute to open evaluation modelling efforts, and to assist in planning and policy development for RTI. Obviously, this is an original initiative still in early stages of design and development, for which researchers' cooperation in all areas will be highly appreciated. Access and contributions are welcome and made possible at <dtostes@gmail.com>.

Further efforts of π-Tupi's development team seek understanding and mechanisms for building complementarity between this Brazilian open access and evaluation's facilitation embryonic endeavour and others internationally gaining momentum, organization and recognition whether in ethical or technical, practical terms (see Figure 2). It is very important to notice that π-Tupi is not aimed at substituting any initiative already in place, but could actually contribute with a different perspective and maybe work integrated to some research centers or RTI social network platforms, enhancing the features provided to their users with less computational resources being needed.

REFERENCES

Pandelis Perakakis. Report back from the COAR 2016 annual meeting. http://www.openscholar.org.uk/report-back-from-the-coar-2016-annual-meeting-2/.

Nicholas, David, et al. „New ways of building, showcasing, and measuring scholarly reputation." Learned Publishing 28.3 (2015): 169-183.

Kreiman, G., & Maunsell, J. H. (2011). Nine criteria for a measure of scientific output. Beyond open access: visions for open evaluation of scientific papers by post-publication peer review, 11.

Samuel, G. N., & Derrick, G. E. (2015). Societal impact evaluation: Exploring evaluator perceptions of the characterization of impact under the REF2014. Research evaluation, rvv007

Thelwall, M., & Kousha, K. (2014). Academia. edu: social network or academic network?. Journal of the Association for Information Science and Technology, 65(4), 721-731.

Thelwall, M., & Kousha, K. (2015). ResearchGate: Disseminating, communicating, and measuring Scholarship?. Journal of the Association for Information Science and Technology, 66(5), 876-889.

Bohannon, J. (2016). The frustrated science student behind Sci-Hub. Science, 352(6285), 511-511.

SWAN, Alma. Why Open Access for Brazil? Liinc em Revista, v. 4, n.2, set 2008.Albagli, S. (2013). Fronteiras da Ciência da Informação.

Maciel, M. L., Abdo, A. H., & Albagli, S. (2015). Open Science, open issues.

Gomes Tostes, Diego et al. Plataforma P2p Para Depósito de Artigos Científicos. VI Confoa. https://repositorio.ufba.br/ri/bitstream/ri/18066/1/CONFOA-2015-folder-virtual.pdf

AUTHORS

DIEGO GOMES TOSTES
Rebec/Fiocruz

PROFA DRA. LUIZA ROSÂNGELA DA SILVA
Rebec/Icict/Fiocruz; FCS/UERJ

PROF. DR. JOSUÉ LAGUARDIA
Icict/Fiocruz

JULIANA GONÇALVES DOS REIS
Rebec/Fiocruz

ELIAS RODRIGUES DE OLIVEIRA FILHO
Rebec/Fiocruz

ROBERTA LOUREIRO BARDUSCO
Rebec/Fiocruz

MARCELLY MACHADO
Rebec/Fiocruz

ANNEX 1

FIGURE 3
Experimental criteria used to assess ongoing intiatives in OA in relation to the proposed solutions

	Sci-hub	Academia edu	Research gate	Repositorio temático	Repositorio institucional	Revistas	π-Tupi
Open source	NO	NO	NO	Depends on the case	Depends on the case	Depends on the case	YES
P2P transfer	NO	NO	NO	NO	NO	NO	YES
Community	NO	YES	YES	YES	NO	Depends on the case	YES
Membership institution	NO	YES	YES	Depends on the case	NO	YES	YES
Membership is independent of knowledge area	NO	YES	YES	NO	Depends on the case	YES	YES
Use is independent	YES	NO	NO	NO	Depends on the case	YES	YES
Altmetrics / quantitative	YES	YES	YES	YES	YES	NO	YES
Webometrics (link analysis)	Depends on the case	YES	YES	YES	YES	NO	YES
Transparent identification	NO	YES	YES	YES	YES	YES	YES
Optional transparent identification	YES	Depends on the case	Depends on the case	Depends on the case	NO	NO	YES
Altmetrics / reputation attributed among members	NO	YES	YES	Depends on the case	NO	Depends on the case	YES
Altmetrics / reputation of the content attributed by those deposits	NO	YES	YES	NO	NO	Depends on the case	YES
Altmetrics / reputation of the content attributed by who retrieves	NO	YES	YES	NO	NO	Depends on the case	YES
Comments, ranking and observations	NO	YES	YES	Depends on the case	Depends on the case	Depends on the case	YES
Access to files outside of the bases of publishers and universities	Depends on the case	Depends on the case	Depends on the case	NO	NO	NO	YES
Automated access to file depends on the base or does not exist	NO	YES	YES	YES	YES	Depends on the case	YES
Private funding	NO	YES	YES	Depends on the case	Depends on the case	Depends on the case	NO
Unknown funding	YES	NO	NO	Depends on the case	Depends on the case	Depends on the case	NO
Mixed funding	NO	NO	NO	Depends on the case	Depends on the case	Depends on the case	NO
Public funding	NO	NO	NO	Depends on the case	Depends on the case	Depends on the case	YES
Paid by the general public	NO	NO	NO	NO	NO	Depends on the case	NO
Paid by the reader / retriever	NO	NO	NO	NO	NO	Depends on the case	NO
Paid by the author / depositor	NO	NO	NO	NO	NO	Depends on the case	NO

White areas meaning NO

darkest areas meaning YES.

Pale grey meaning "depends on the case"

Survey performed by the authors, 2015/2016

THE USE OF SYNERGIES BETWEEN RESEARCH FUNDING AND CRIS SYSTEMS FOR THE DOCUMENTATION AND EVALUATION OF THE SOCIETAL IMPACT OF APPLIED AGRICULTURAL RESEARCH

BIRGE WOLF AND JÜRGEN HESS

CURRENT SITUATION AND PROBLEM

Publicly financed agricultural research should be scientifically excellent, societally relevant and, to a substantial extent, geared to practical implementation (EC SCAR 2014). In the scientific system, however, it is mainly publications and high citation rates that are acknowledged, and this in turn creates incentives within the system to pursue those very aims.

Accordingly, ways of increasing and acknowledging the societal impact of research are being discussed in many research fields, and a broad range of concepts have been developed for evaluation related to societal impact (Penfield et al. 2014, Wolf et al. 2013). Examples are the focus on quality assurance in transdisciplinary research, institutional evaluation such as the societal impact assessment of agricultural research conducted by the French INRA (Joly et al. 2015), and national assessments in the UK and NL. In the UK's Research Excellence Framework (REF2014), the performance-based allocation of 20% of the funds received by higher education institutions is geared to societal impact (HEFCE 2014). In the Netherlands, societal relevance is part of the institutional statement on research outcomes; evaluation results are used for institutional development, but not for resource allocation (VSNU et al. 2014). A number of research funders focus in their programmes and funding decisions on expected societal impacts and measures taken to achieve them, (e.g. EC 2014), but programme evaluations do not usually acknowledge projects' commitment to societal benefit.

Thus our work starts with the question of how an evaluation of the societal impact of agricultural research can be introduced in Germany. National approaches do not seem viable as a starting point, due to federalist structures. We therefore concentrate on synergies with research funders, because those who focus on societal impact are interested in maximising and demonstrating it. Besides the classical impact assessment challenges such as time gap, temporality of impact, attribution gap and impact evidence, the often described lack of adequate, easy-to-use data for evaluation is an important point: funders require the documentation of activities and outputs related to their specific programme goals, but they often want them described in a more concrete and complete way, and wish to be informed about outputs and impacts after conclusion of the project.

Improvements of documentation facilities related to societal impact are the subject of constant, dynamic development. Research Councils UK collects outputs and outcomes in a period up to three years after project completion with 'researchfish' (formerly the 'Research Outcome System'). Work is also being done on the integration of information on the societal outputs of research in current research information systems (CRIS) (Jörg et al. 2014, Gartner et al. 2013), with the focus on the indicators of the REF and Research Councils UK.

Thus, based on these developments, we focus on how documentation and evaluation of research contributions to societal impact can provide even more benefits for research funders, so that they can contribute to changes in scientific incentive systems towards societal impact. In principle, impact assessments have to consider the time lag between research and impact, the change of impact over time and the complexity of innovation systems with various actors and framework conditions beyond science (Penfield et al. 2014, Spaapen and van Drooge 2011). In order to acknowledge contributions of research to societal impact in a time frame that is relevant for institutional or personal development, our concept focuses on explicit, substantial contributions of research rather than on the attribution and evidence of its real impacts as in REF2014.

OBJECTIVES

The overarching aim of the project is to enhance incentives for societally relevant research in Germany taking the example of agricultural research.

To contribute to this aim, the project Praxis Impact II[1] has drawn up an extended documentation and evaluation concept for applied agricultural research which provides synergies with research funding. These

1 Supported by the German Federal Ministry of Food and Agriculture following a decision by the German Federal Parliament.

synergies aim to a) increase acknowledgement of research oriented toward practice and society, independent of whether national assessment procedures exist or not, b) assist funders in augmenting the societal benefit of funded research, and c) reduce the amounts of time and work required for documentation and evaluation.

MATERIAL AND METHODS

1. A literature review of evaluation concepts (societal / broader impact, interdisciplinary and transdisciplinary research, evaluation of development cooperation) was undertaken. Additionally, qualitative interviews were conducted with researchers from applied and transdisciplinary research in organic agriculture and adjacent research areas (results in Wolf et al. 2013). Furthermore, tools and standards related to societal impact documentation within CRIS (e.g. Jörg et al. 2014, CASRAI 2014, euroCRIS 2013), requirements for project application, and information about federal research funding in Germany were reviewed. The different approaches in the literature were used to develop the documentation and evaluation concept.
2. After preparatory work in spreadsheets, the documentation concept was embodied as a prototype in a database with a user interface that enables project information related to societal impact and project management aspects to be documented in a way that fits into the logical framework of proposals and reports.
3. Iterative test cycles of the documentation concept were conducted with twelve projects, focusing on data input in the context of project application and reporting by researchers. The prototype was improved according to the test results and subsequently used to conduct further test cycles. In the first cycle information was entered by our project team, in the second by other researchers. In order to record problems and opportunities for improvement, the ‚think-aloud method' (Häder 2010) was used directly during the processing of the entry forms. Individual aspects were looked at in more detail in dialogue and all the statements made by the test subjects were recorded by the project team.
4. Three workshops were conducted to discuss the approach with the funding agency.
5. The information on three projects, gathered in the data input test cycles, was used to compile reports for the evaluation of contributions to societal impact. The evaluation tests were conducted by consultants and researchers with expert knowledge in the fields concerned by means of the evaluation scheme developed. The tests focused on the usability and quality of the evaluation guidelines and the suitability of the information gathered with the prototype.
6. The records of all the tests were encoded, and measures for adaptation were derived from the results and implemented.

RESULTS AND DISCUSSION

EVALUATION PROCEDURE

The evaluation concept was developed in iterative cycles with the documentation concept. It serves to evaluate individual projects or groups of thematically related projects two to three years after their conclusion. The information, updated by those carrying out the project, is supplemented with feedback from practice and society. It is then reviewed by scientists and representatives from practice and relevant areas of society. With a list of questions, experts are guided through the evaluation, the aim being for it to end up in a summarised classification of a project into one of 5-6 different quality stages. The evaluation process should be organised and supervised by evaluation experts and financed by the funder.

CRITERIA

To compile the criteria and information needed for evaluation (Fig. 1) we take into account the diverse nature of impact pathways in agricultural research. These can include linear technology transfer as well as transdisciplinary approaches. These diverse criteria that may indicate a contribution to societal impact are clustered into three guiding questions for evaluation. The first compiles information that is needed to estimate whether research deals with societally relevant issues and conducts relevant processes that enhance the focus on applicability and societal benefits, for example via interactions with stakeholders. The second focuses on applicable outputs and their suitability and accessibility for target groups. The third covers a broad range of possible or de facto applications including changed understanding etc., innovations in services, products and processes, social innovation, influence on policy etc., and spin-offs and intellectual property rights. Participants in testing procedures and workshops have acknowledged the comprehensiveness of the documentation approach.

Fig. 1: Criteria of the evaluation concept provided with diverse information from the documentation concept

Thus, in our view, a contribution to societal impact requires researchers to make a commitment to applicability for non-academic actors and to (potential) societal benefits (in terms of sustainable development) that are associated with an application. To cater to the complexity of innovation systems, both information on relevant stakeholders and target groups and descriptions of framework conditions (e.g. market development, state support, legal restrictions, public awareness) are part of our concept. Furthermore, the evaluation concept makes provision for the independent recording of feedback from and dialogue with actors from practice and society, for different reasons. Firstly, to increase the plausibility of information documented by researchers with a reasonable amount of effort, as recommended for the evaluation of research for development (Saint-Martin et al. 2012, p. 6), because experience in REF2014 shows that providing evidence of the societal impacts of research is viewed by researchers as involving a great deal of time and work (Manville et al. 2015). Secondly, when it is a question of complex innovation processes and assessing the soundness of possibilities for application, the triangulation of perspectives and the focus on plausibility may be more appropriate than the focus on evidence, because in REF2014 there was shown

1 Supported by the German Federal Ministry of Food and Agriculture following a decision by the German Federal Parliament.

to be a preference among evaluators for acknowledging quantitative and easily assessable evidence (Vertigo-Ventures and Digital Science 2016: 12). In the impact case studies in medicine, furthermore, linear impact pathways were shown to be dominant (Greenhalgh and Fahy 2015). Thirdly, stakeholder involvement is suitable for providing learning and enhancing the conditions for current and further innovation processes (Coutts et al. 2014, Ruane 2014), which serves funders' programme goals and may also encourage researchers.

USE OF APPLICATION AND REPORTING PROCEDURES

To provide synergies with research funding processes, we developed and iteratively tested the user interfaces that collect information related to contributions to societal impact within application and reporting procedures. This led to the following findings:

a. A fundamental reduction of documentation redundancies is possible if categories of contributions to societal impact are introduced directly at the point where the aims of a project are formulated. This means that instead of being formulated in a single text, aims are documented directly in suitable categories such as activities, publications, products, applicability, intended impact etc. These categorised aims may support funding decisions and project administration and can be used and edited for reporting.
b. The description of the work packages is supported in our documentation concept as follows. Firstly, intended outputs and outcomes from the aim description can be linked to certain work packages and this reduces the need to describe them again as part of the work conducted. Furthermore, we included functions to facilitate time planning in the work packages that were suggested by scientists in order to increase work efficiency. Thus, even if this is not crucial to societal impact assessment, the tests do indicate that it supports the acceptance and usability of the documentation system.

CONCLUSIONS AND PERSPECTIVES

The documentation concept and evaluation questions have been tested with scientists and consultants, but so far only discussed with representatives of research funders. Meaningful next steps towards a system that could actually be implemented would therefore be a) detailed test processes in order to integrate the requirements of the funder more firmly; b) the extension of a professional CRIS based on the documentation concept; c) further tests of documentation and evaluation with all the users prior to and during gradual implementation. The following benefits can be achieved with the implementation of the concept thus developed:

a. A documentation system which replaces parts of text-based applications and reports obviates additional documentation work and, combined with an evaluation procedure, provides sufficient incentive to deliver multiple societal outputs and document them in their entirety. With this substantial data basis, a well founded project evaluation can be implemented with a reasonable amount of effort.
b. The evaluation findings are useful to the funding providers and can increase the incentives for practice- and society-oriented research in the scientific system. 'Award-winning projects' (weighted with the amount of funding) may also be an easy-to-use indicator in the evaluation of institutes or scientists.
c. The information in the system supports target-oriented programme implementation, proposal reviews and project supervision.
d. Knowledge and technology transfer are increased, increased, because not only scientific publications, but also outputs and exchange formats for target groups in practice and society can be archived consistently, continued after project completion and made accessible both immediately and in the long term.

REFERENCES

CASRAI (2014): „Homepage, Dictionary," retrieved on 06 May 2016 from http://dictionary.casrai.org

EC (European Commission) (2014): „Guidance for evaluators of Horizon 2020 proposals," retrieved on 06 May 2016 from http://ec.europa.eu/research/participants /data/ref/h2020/grants_manual/pse/h2020-evaluation-faq_en.pdf

EC SCAR (European Commission, Standing Committee on Agricultural Research) (2014): „Agricultural knowledge and innovation systems towards 2020," Publications Office, Luxembourg.
euroCRIS (2013): „cerif release 1.6," retrieved on 06 May 2016 from http://www.eurocris.org/cerif/main-features-cerif

Gartner, R., M. Cox, K. Jeffery (2013): „A CERIFÐbased schema for recording research impact," The Electronic Library 31, 4/2013, pp. 465–482.

Greenhalgh, T., N. Fahy (2015): „Research impact in the community-based health sciences: an analysis of 162 case studies from the 2014 UK Research Excellence Framework," BMC medicine 13, 2015, p. 232.

Häder, M. (2010): „Empirische Sozialforschung. Eine Einführung," VS, Verl. für Sozialwiss., Wiesbaden.

HEFCE (Higher Education and Funding Council for England) (2014): „REF 2014 Key facts," retrieved on 06 May 2016 from http://www.ref.ac.uk/media/ref/content/ pub/REF%20Brief%20Guide%202014.pdf from.

Joly, P.-B., A. Gaunand, L. Colinet, P. Larédo, S. Lemarié, M. Matt (2015): „ASIRPA: A comprehensive theory-based approach to assessing the societal impacts of a research organization," Research Evaluation 24, 4/2015, pp. 440–453.

Jörg, B., S. Waddington, R. Jones, S. Trowell (2014): „Harmonising Research Reporting in the UK – Experiences and Outputs from UKRISS," Procedia Computer Science 33, 2014, pp. 207–214.

Manville, C., M. Morgan Jones, M. Frearson, S. Castle-Clarke, M.-L. Henham, S. Gunashekar, G. Jonathan (2015): „Preparing impact submissions for REF 2014: An Evaluation. Findings and observations," retrieved on 06 May 2016 from http://www.rand.org/content/dam/rand/pubs/research_reports/RR700/RR727/RAND_RR727.pdf

Penfield, T., M. J. Baker, R. Scoble, M. C. Wykes (2014): „Assessment, evaluations, and definitions of research impact: A review. Research Evaluation 23, 1/2014, pp. 21–32.

Spaapen, J., L. van Drooge (2011): „Introducing ‚productive interactions' in social impact assessment. Research Evaluation 20, 3/2011, pp. 211–218.

Vertigo-Ventures; Digital Science (2016): „Collecting Research Impact Evidence. Best Practice Guidance for the Research Community", retrieved on 28 June 2016 from http://www.vertigoventures.com/research-impact-evidence-download

VSNU, NWO, KNAW (2014): „Standard Evaluation Protocol 2015 – 2021," retrieved on 06 May 2016 from https://www.knaw.nl/shared/resources/actueel/publicaties/pdf/ standard-evaluation-protocol-2015-2013-2021

Wolf, B., T. Lindenthal, M. Szerencsits, J. B. Holbrook, J. Heß (2013): „Evaluating Research beyond Scientific Impact – How to Include Criteria for Productive Interactions and Impact on Practice and Society. GAIA - Ecological Perspectives for Science and Society 22, 2/2013, pp. 104–114.

AUTHORS

BIRGE WOLF
University of Kassel, Organic Farming and Cropping Systems
Nordbahnhofstr. 1a,
37213 Witzenhausen
Germany
+49 5542 981536
birge.wolf@uni-kassel.de

JÜRGEN HESS
University of Kassel, Organic Farming and Cropping Systems
Nordbahnhofstr. 1a,
37213 Witzenhausen
Germany
+49 5542 981536
birge.wolf@uni-kassel.de

AUSTRIAN PLATFORM FOR RESEARCH AND TECHNOLOGY POLICY EVALUATION (fteval)
c/o ZSI – Centre for Social Innovation GmbH
Linke Wienzeile 246, A-1150 Vienna
T +43 1 495 04 42 - 79
F +43 1 495 04 42 - 40
E office@fteval.at
W www.fteval.at

DESIGN
carotte.at
Blindengasse 35/4/97, A-1080 Wien
E caroline@carotte.at

PRINT
REMA-Print-Littura Druck- u. Verlags Ges.m.b.H
Neulerchenfelderstraße 35, A-1160 Vienna

AUSTRIAN PLATFORM FOR RESEARCH AND TECHNOLOGY POLICY EVALUATION
Platform fteval members:

Federal Ministry of Science, Research and Economy (bmwfw), Federal Ministry of Transport, Innovation and Technology (bmvit), Austrian Institute of Technology (AIT), AQ Austria, Austria Wirtschaftsservice (AWS), Christian Doppler Research Association (CDG), convelop cooperative knowledge, Austrian Research Promotion Agency (FFG), Austrian Science Fund (FWF), Institute for Advanced Studies (IHS), Austrian Institute for SME Research (KMFA), Joanneum Research, Ludwig Boltzmann Society (LBG), Austrian Institute of Economic Research (WIFO), Austrian Council for Research and Technology Development, Austrian Academy of Science (ÖAW), Technopolis Austria, Vienna Science and Technology Fund (WWTF), Wirtschaftsagentur Wien. Ein Fonds der Stadt Wien, Centre for Social Innovation (ZSI)

AUSTRIAN PLATFORM
for Research and Technology Policy Evaluation

fteval

c/o ZSI – Centre for Social Innovation GmbH
Linke Wienzeile 246, A-1150 Vienna

T +43 1 495 04 42 - 79
F +43 1 495 04 42 - 40
E office@fteval.at
W www.fteval.at

The fteval Journal for Research and Technology Policy Evaluation is a forum to discuss current evaluation practices in the field of research, technology and innovation policy.

ZVR-Zahl: 937261837
ISSN-Nr. 1726-6629