39

fteval JOURNAL for Research and Technology Policy Evaluation

ISSUE 39 | SEPTEMBER 2014

CONFERENCE
»Evaluation of STI policies, instruments and organisations: New horizons and new challenges, Vienna 2013«

EMANUELA REALE, MARIA NEDEVA, THOMAS DUNCAN & EMILIA PRIMERI

EVALUATION THROUGH IMPACT: A DIFFERENT VIEWPOINT

MATHIAS WEBER & WOLFGANG POLT

ASSESSING MISSION-ORIENTATED R&D PROGRAMS: COMBINING FORESIGHT AND EVALUATION

CLAUDIA MICHEL, SIMON HEARN, GABRIELA WUELSER & THOMAS BREU

ASSESSING THE BROAD SOCIETAL IMPACTS OF RESEARCH: THE CASE OF THE NCCR NORTH-SOUTH PROGRAMME

FEDERAL MINISTRY FOR SCIENCE, RESEARCH AND ECONOMY

Minoritenplatz 5, A -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, A-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 Annina Müller-Strassnig
E: annina.mueller@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
DI 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, A-1090 Vienna
DI Dr.in Sabine Mayer
E: sabine.mayer@ffg.at
Mag. Leonhard Jörg
E: leonhard.joerg@ffg.at

FWF - AUSTRIAN SCIENCE FUND

Sensengasse 1, A-1090 Vienna
Dr. Rudolf Novak
E: novak@fwf.ac.at
Dr.in Dorothea Sturn
E: dorothea.sturn@fwf.ac.at

IHS - INSTITUTE FOR ADVANCED STUDIES

Stumpergasse 56, 1060 Vienna
Mag. Richard Sellner
E: richard.sellner@ihs.ac.at

JOANNEUM RESEARCH FORSCHUNGSGESELLSCHAFT MBH

Sensengasse 1, A-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, A-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

TECHNOLPOLIS AUSTRIA

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

WIFO - AUSTRIAN INSTITUTE OF ECONOMIC RESEARCH

Arsenal, Objekt 20, PF 91,
A-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, A-1090 Vienna
Dr. Michael Stampfer
E: michael.stampfer@wwtf.at
Dr. Michael Strassnig
E: michael.strassnig@wwtf.at

ZIT - THE TECHNOLOGY AGENCY OF THE CITY OF VIENNA

Ebendorferstraße 4/DG,
A-1010 Vienna
Robert Mayer-Unterholzer
E: robert.mayer@zit.co.at

ZSI - CENTRE FOR SOCIAL INNOVATION

Linke Wienzeile 246, A-1150 Vienna
Dr. Klaus Schuch
E: schuch@zsi.at
Mag.ª Elke Dall
E: dall@zsi.at

TABLE OF CONTENTS

ISSUE 39 | 2014

PREFACE

KLAUS SCHUCH, managing director fteval

DEAR READER!

fteval Journal for Research and Technology Policy Evaluation number 39, the journal in hand, is a special issue; not only a heavy one.

It includes 16 papers, some more academic, some more practice orientated, as well as a couple of short session summaries from the conference **„New Horizons / New Challenges: evaluation of STI policies, instruments and organisations"**, which was organised by the *Austrian Platform for Research and Technology Policy Evaluation (fteval)*, *the Manchester Institute of Innovation Research (MIOIR) and L'IFRIS - institut Francilien Recherche Innovation Société,* in Vienna in November 2013.

This international conference provided an open forum to debate developments in science, technology and innovation (STI) policy and their effects on evaluation theory and practice.

More than 200 evaluators, scientists, research managers, authorities and STI policy makers from 28 countries, including the non-EU countries Kazakhstan, Korea, Serbia, Russia, Thailand, Ukraine, Uruguay and the USA participated in the conference. They discussed the next generation of STI policy evaluation designs and approaches, which are challenged by different developments such as the emergence of new demand-side oriented instruments, the increasing complexity of appraisals, the demand for impact assessments beyond economic effects measurement, an increasing amalgamation between national and European/international interventions in STI or the emphasis on new mission-oriented approaches ("grand challenges"), just to name a few.

This fteval journal includes papers from any of the ten conference sessions.

Mathias Weber from the Austrian Institute of Technology and **Wolfgang Polt** from Joanneum Research contributed to session 1 "New approaches for evaluating STI Policies and Instruments" with a paper on *"Assessing mission-orientated R&D programs: combining foresight and evaluation"*.

For the session 2 "Assessing the variety and long-term impact of research" we are delighted to have two papers published in this issue; one by **Federica Rossi** from Birbeck University of London, **Annalisa Caloffi** from the University of Padova and **Margherita Russo**, University of Modena and Reggio Emilia, on *"Can policy design help organizations improve their networking capabilities? An empirical analysis on a regional policy"*; and one by **Claudia Michel**, University of Bern, **Simon Hearn**, Overseas Development Institute ODI, **Gabriela Wuelser**, Swiss Federal Institute of Technology Zurich and **Thomas Breu**, University of Bern, on *"Assessing the broad societal impacts of research: the case of the NCCR North-South programme"*.

Session 3 about "STI Policy Evaluation in new- and non-OECD Countries" is also represented by two papers from **Vitalii Gryga** and **Victor Rybachuk**, both from the STEPS Center of the National Academy of Sciences of Ukraine and **Olha Krasovska**, State Fund for Fundamental Research at the State Agency of Ukraine of Science, Innovation and Information on *"Evaluation of R&D Institutions in Ukraine – The New Approach"*; and from **Yuri Simachev**, **Mikhail Kuzyk** and **Vera Feygina**, all from the Interdepartmental Analytical Center, on *"Promoting firms' innovative behavior in Russia: what weakens the power of STI policy?"*.

Also from session 4 "Challenges in assessing new European Research Area polices, programs and instruments " two interesting papers are included in this issue: one from **Emanuela Reale**, CERIS CNR, **Maria Nedeva** and **Thomas Duncan** from University of Manchester/ Manchester Institute of Innovation Research, and **Emilia Primeri**, also CERIS CNR, on *"Evaluation through impact: a different viewpoint"* and one from **Martin Felix Gajdusek**, Centre for Social Innovation, and **Nikos Sidiropoulos**, University of Athens, on *"Monitoring and Evaluation in joint calls of "horizontal – INCO" ERA-NET and ERA-NET PLUS actions"*.

Susanne Bührer from the Fraunhofer Institute for Systems and Innovation Research contributed to session 5 "Evaluating for selection – challenges and opportunities" with a paper on *"New modes of stakeholder involvement in ex-ante impact assessments"*.

Session 6 on "Evaluation practices scrutinized" is represented in this issue by two contributions: one from **Erich Prem** from eutema Technology Management, on *"Evaluation as the Construction of Policy Narratives"*; and one from **Franz Barjak**, University of Applied Sciences and Arts Northwestern Switzerland, on *"Wie wirksam sind Innovationsfördermassnahmen in der Schweiz?"*.

Christiane Kerlen from Dr Kerlen Evaluation, **Jan Wessels**, Institut fur Innovation und Technik, and **Volker Wiedmer**, Hochschule Magdeburg-Stendal, represent conference session 7 "Evaluation of STI policy portfolios and policy mixes" with a contribution on *"Portfolio evaluation: A case study to illustrate evaluation challenges"*.

Two papers published in this issue were presented during conference session 8 "Data, monitoring systems and indicators": one from **Matteo Razzanelli**, Science Europe, on *"The European Research Area: a process-related challenge for indicator and policy design"* and one by **Michael Dinges**, Austrian Institute of Technology, **Jakob Edler**, University of Manchester/ Manchester Institute of Innovation Research and **Matthias Weber**, also Austrian Institute of Technology, on *"Positioning competence centres via monitoring data: towards a systematic approach"*.

Sonja Kind and **Gerd Meier zu Köcker**, both from the Institute for Innovation + Technology represent session 9 "New Developments in Innovation Policy Evaluation" with a contribution on *"Evaluation of Clusters, Networks and Cluster Policies – Challenges and Implementation"*.

Session 10 on "Evaluation of International RTI Programmes" is finally represented in this issue with a contribution of **Christina Schuh** from the Humboldt Foundation on *"Expectations on the long-term impact of research fellowships from an evaluation perspective: challenges and limits to measure side-effects"*.

We are sincerely grateful to all authors who devoted their time and energy in providing us with a full paper based on their conference presentations!

Experimentally, we also include in this fteval journal issue a few short conference summaries drafted by students from the Department of Science and Technology Studies (S&TS) of the University of Vienna. It is interesting to retrace their impressions, reflections and perceptions. In addition, we included also two more comprehensive panel summaries drafted by **Mario Steyer** (bmvit – Austrian Federal Ministry of Transport, Innovation and Technology) and by **Martina Lindorfer** (ZSI – Centre for Social Innovation). Thank you very much for that!

Our next international conference is not yet fixed, but first ideas are looming. In the meanwhile I hope you find a contribution in this issue of the fteval Journal for Research and Technology Policy Evaluation which attracts your attention.

Enjoy reading!

Klaus Schuch
fteval

ASSESSING MISSION-ORIENTATED R&D PROGRAMS: COMBINING FORESIGHT AND EVALUATION

K. MATTHIAS WEBER, WOLFGANG POLT

ABSTRACT

In parallel with the rise of policy interest in major societal and environmental challenges since the 1990ies, reflected e.g. in the Millennium Development Goals and the forthcoming Sustainable Development Goals, R&D - and publicly funded R&D in particular - is increasingly expected to contribute to resolving such challenges. This new kind of 'mission-orientation' has developed into an important rationale for legitimizing public R&D spending in times of budgetary constraints and tighter requirements of setting priorities for public spending. Such re-emphasis is to be seen both at the level of national (e.g. in the German High-Tech-Strategy) as well as at the international level (e.g. in the Horizon 2020 Framework Programme of the European Union).

However, it is also a rationale that is very difficult to pin down in economic terms as the expected benefits often lie quite far in the future, and sometimes need the transformation of entire production-consumption systems to be fully realized. The impact of R&D policies and programmes cannot be assessed independently of change processes in specific sectors, which determine the likelihood of diffusion of innovations, and ultimately of transformative change processes, and thus only in conjunction with the corresponding sectoral policies.

In such a complex setting, the attribution of impacts to specific RTI policy programmes represents a major difficulty. We propose a methodological and operational framework for the evaluation of mission-oriented programmes that takes these challenges into account. It distinguishes guiding principles for ex-ante assessment and ex-post evaluation and proposes a "PESCA (Prospective & Adaptive Societal Challenges Assessment) Approach", to assessing mission oriented policy. Strong emphasis in this approach is put on ex-ante impact assessment, social cost benefit analysis, an adaptive and flexible process through which assessment results are generated and fed back to policy.

KEY TRENDS IN RTI POLICY: EMPHASIZING NEW MISSION-ORIENTED RTI PROGRAMMES

In recent years and in an increasing number of countries we could observe a re-newed emphasis on funding initiatives that are supposed to help tackle major, often long-term, societal challenges. This 'new mission-oriented policy' differs in several aspects form its earlier predecessor (e.g. the post-war policies focusing on large-scale initiatives in aerospace, defence, or energy) but also from the more 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". While these "new" mission-oriented R&D programs may share an explicit thematic goal-orientation with the "old" mission-oriented programs of the Sixties, they are not solely guided by technological, but pre-dominantly by societal targets.[1] The emergence of new mission-oriented R&D programmes is not accompanied by a demise of structure-oriented programmes, but rather builds on the performance of well-established national innovation systems for enabling goal-oriented change for tackling societal challenges.

With this change of the purpose of R&D, the requirements for their evaluation have equally changed. To this adds a stronger emphasis put by public authorities on the ex-ante assessment of expected or likely impacts of policy initiatives, which complements the by now well-established procedures of ex-post evaluation.

This new situation and the new characteristics of this type of policies raise a number of fundamental challenges for (ex-ante) impact assessment and subsequent (ex-post) impact evaluation. This paper aims to elaborate these challenges in a systematic manner. It also outlines a systematic approach and guiding principles for conducting ex-ante impact assessment and ex-post evaluations of new mission-oriented programmes, and points at a number of ongoing programmes where these insights could be fruitfully brought to bear.

1 For an overview of historical shifts in RTI policy, see Gassler, Polt and Rammer (2008).

CHALLENGES AND REQUIREMENTS FOR THE ASSESSMENT AND EVALUATION OF MISSION-ORIENTED RTI PROGRAMMES

There are some characteristics of 'new mission-oriented policies/programmes (MOPs)' which have emerged in the past couple of years and which define the requirements for their assessment:[2]

- Most recent MOPs – corresponding to the nature of societal challenges – are addressing issues that are broader in nature and scope than earlier technology-centred 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. Contrary to old MOPs, their most recent variants would ascribe a much larger role to private sector actors.

- 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 programmes 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 programmes like the often-cited role model of the earlier types of MOPs (e.g. the Manhattan and the Apollo programmes) even programmes confined to one field or topic (e.g. the US energy programmes) 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.

In short, today's MOPs can be interpreted as 'systemic policies in a nutshell' (though they might be large nuts!) with most of the characteristics and obstacles systemic policies face in general.

While typical commercial, micro-level effects can be analysed with the help of well-established assessment and evaluation methods, this systemic policy approach 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[3] before it can actually exert an influence on new mission goals. The immediate impact of a mission-oriented R&I programme occurs at the level of the participating firms or research organisations, where new research results are produced and – at least in some cases – innovations are introduced to the market.

However, it is only after widespread uptake and diffusion of an innovation in the target system that an impact of a mission-oriented R&I programme 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, changes are also needed at different levels of the target systems. Borrowing from the multi-level perspective on socio-technical transitions,[4] 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). The dominant socio-technical regime, however, raises important constraints for a potential transition of the target system and for the potential mission-oriented impacts to be induced by RTI policy programmes.

Most "new missions" as the guiding aims of funding programmes tend to be defined at the level of such meso-level socio-technical regimes. Realizing these missions requires the widespread uptake and diffusion of innovations, if not a transformation of the production and consumption practices.

METHODOLOGICAL IMPLICATIONS FOR EX-ANTE IMPACT ASSESSMENT AND EX-POST EVALUATION: TOWARDS A PROCESS MODEL

LEVELS AND PATHWAYS OF IMPACT: A FRAMEWORK

Against the background of the above characteristics and requirements of MOPs, the subsequent section aims to outline a novel me-

2 See for an earlier description Soete and Arundel (1993) and for a more recent one Foray, Mowery and Nelson (2012) which shape the requirements of their assessments. It has to be added, though, that mission-oriented policies can have different characteristics in terms of goals, instruments, stakeholders and effects. Hence, the characteristics given do not apply to the same extent and at the same time to all types of mission-oriented programmes.

3 We conceptualized four different stages (see below).

4 Here we have adopted the three levels or domains of analysis that have been suggested by the transitions literature (see Geels 2002).

thodological framework for the evaluation and assessment of MOPs. Established R&I programme evaluation methodologies focus mainly on the impact of funding programmes at the level of niches, with a view to the increase in innovation performance and research outputs, but tend to restrict the impact analysis at regime level to economic matters such as competitiveness and employment, or an outlook on technological or at best techno-economic potentials of the supported R&I activities. Some programme evaluations with a dedicated diffusion-orientation have focused on the uptake of new technologies in industry, as well as organisational implications they have raised.[5] Others have attempted to demonstrate impacts on employment. However, these approaches capture only some aspects of what is understood nowadays by societal missions.

We therefore propose a conceptual framework to underpin the study of impacts which builds on two main dimensions:

- First, the '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.
- Second, the 'impact level': Contributing to the achievement of mission goals implies changes to be realized at different levels, i.e. changes at micro-level of individual behaviour, 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.

This simple framework implies that rather than looking at innovation systems in the traditional sense, we need to study impacts on mission goals within a framework of "systems of innovation, production and consumption [SIPC]" (Weber and Rohracher 2012) as the frame of reference. While maintaining a systems language, this perspective draws much broader boundaries for system analysis, impact assessment and evaluation than the traditional innovation systems perspective. It also looks at the interdependencies between innovation activities on the one hand and production-consumption practices on the other hand. One could argue that SIPC integrates two hitherto separate streams of system analysis, namely innovation systems analysis and the analysis of production-consumption systems (e.g. Tukker et al. 2008)

What needs to be explored for purposes of impact assessment and evaluation against the background of this broadened view on innovation, production and consumption are impact pathways that are non-linear and often involve feedback and rebound mechanisms between the levels and/or phases.

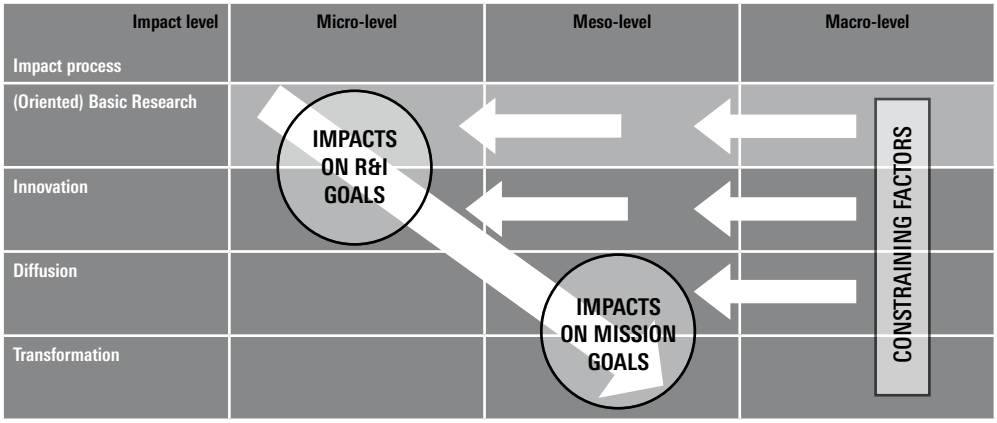


FIGURE 1 - Conceptual framework of impact processes and impact levels

METHODOLOGICAL IMPLICATIONS – EX ANTE

The framework depicted above allows explaining what kinds of impacts should be taken into account in impact assessment, as well as in evaluations, if mission-orientation is taken seriously as a policy target. Adopting it has several methodological implications:

First, we would see a shift in emphasis towards ex-ante assessments. Policy frameworks with potentially long gestation periods and substantial impacts on societies and economies need to undergo a careful ex-ante analysis of their potential impacts (including unintended ones). For such ex-ante impact assessment, we suggest a forward-looking, scenario-based approach, exploring scenarios at three different levels in order to cope with different types of future contingencies:

- *Context scenarios* to cope with broader contingencies and constraints at the level of a socio-technical landscape. They thus provide different frames and assumptions for bundles of potential impact pathways.
- *System scenarios* are based on a thorough exploration of possible impact pathways that inter-connect micro- and meso-level developments. The systems under study must have a sufficiently broad scope, similar to systems of innovation, production and consumption, if impacts on mission goals are to be studied. As

5 As an example, see the Austrian evaluations of the FlexCim programmes (Geyer et al. 2000).

missions are expressed not just in terms of innovation, but in terms of actual changes in living, working and producing in society, a broader systemic frame must be chosen, which covers both R&I and sectoral/thematic production-consumption aspects.
The knowledge on which the elaboration of such system scenarios draws is a mix of theoretical insights into the structure and dynamics of systems, exploration of current observable trends and developments at micro and meso level, but also of unexpected developments and wildcards, which require a great deal of creativity to be imagined. This knowledge delivers a structured, but at the same time an open understanding of how a system might evolve in the future. Different degrees of openness and relaxation of assumptions about the continuity of current trends are possible; it is just a matter of making such choices explicit. In the same vein, the process by which this knowledge is created and fed back into the policy process must be open and flexible: in case of new options, technological opportunities or changing societal demands, re-considerations of mission targets must be possible to avoid lock-in.

- *Policy and funding scenarios:* Different packages or even roadmaps of RD&I and sectoral policies need to be assessed and compared in terms of their expected impacts on mission goals against the background of different context scenarios and system scenarios. The impacts of these packages of instruments need to be studied with regard to different target system scenarios. This is necessary because the target system scenarios depend also on other factors of influence than the policy instruments under study. Other actors may exert an influence as well. Not the least, system scenarios need to be compatible with the way the wider context evolves, which is expressed in terms of context scenarios.

An impact assessment of a particular funding programme would thus not be conducted in isolation, but the programme would be seen and assessed as part of a package or portfolio of policy instruments, aiming to shape the target system in the direction of the envisaged mission goals. This is essential because the impact of a specific programme is inter-related with that of other policies and initiatives. In fact, recent MOPs tend to bundle different specific instruments, as recognition of the need to apply policy mixes if mission goals are to be approached. A serious impact assessment would thus need to anticipate possible impact pathways, taking into account the interactions between different policy instruments. Such systemic, multi-instrument intervention logic is essential in order to give justice to the complexity of the transformative processes needed to reach mission goals.

Figure 2 gives an overview of how the process of an ex-ante impact assessment of a MOP could look like. The three levels of scenarios correspond to the three vertical streams, addressing the context of the system of innovation, production and consumption (SIPC) under study, the SIPC itself, and the policy and funding instruments that are currently applied or might be in the future. Ultimately, the process of exploring future impacts in a scenarios framework (Steps 1 to 4) should feed into what could be called a social cost-benefit analysis of policy and funding system scenarios with regard to their suitability to reach mission goals for different consistent context-SIPC scenarios (Step 5).

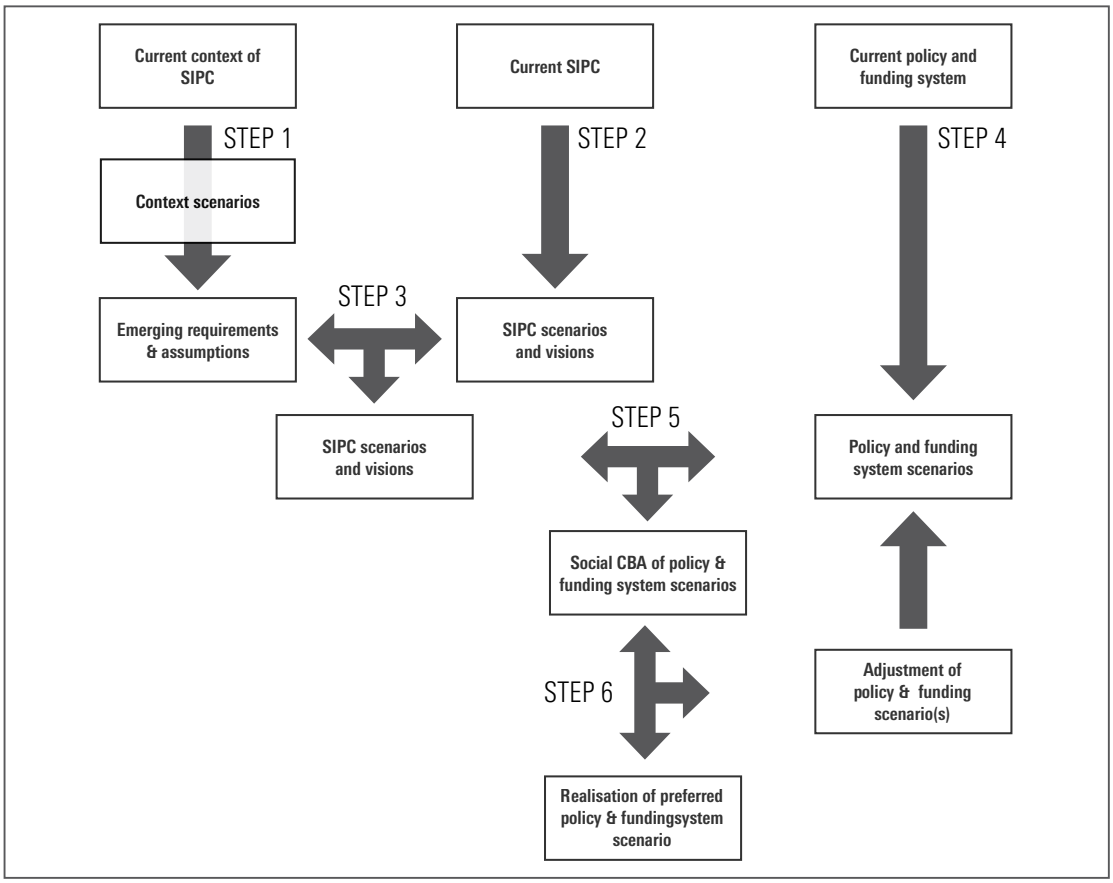


FIGURE 2 - A process model for ex-ante impact assessment of policy instruments on mission goals (based on Weber and Johnston 2008)

However, we should be fully aware of the limits to modelling in quantitative or even monetary terms the kinds of impacts expected. Social costs and benefits need to be understood in qualitative as well as – to the extent possible - quantitative terms. A process of sense-making is thus required that builds conceptually on the notion of social cost-benefit analysis. Depending on complexity of impact pathways, only upper and lower bounds of impacts of MOPs on mission goals can be assessed, while more modest and specific programme goals may be accessible to more precise assessments.

Given the complexity, uncertainty and ambiguity of the future (Renn et al. 2011), it is important to foresee an iterative process of learning (Step 6). The so-called Collingridge dilemma implies that we continuously acquire new knowledge about new social, economic and scientific-technological developments, as well as about the impacts of these developments on mission goals, and – as a consequence – about the impacts of policy instruments (Collingridge 1980). A continuous re-adjustment of policies and instruments is thus of crucial importance for a long-term strategy of new mission-oriented governance.

METHODOLOGICAL IMPLICATIONS FOR EX-POST IMPACT ASSESSMENT

In general, the ex-ante assessment of policy interventions defines also the framework for a subsequent ex-post evaluation. However, when dealing with mission-oriented policy, the evaluative focus naturally has to shift from ex-post to (i) ex-ante impact assessment (especially on ex-ante social cost-benefit assessment) and on (ii) the process of joint vision and policy forming (which is formative by nature!) for a number of reasons:

- Ex-post evaluation of the contribution of the involved RTI policies to the achievement of the mission goal is facing even greater obstacles as evaluation of individual funding programmes because the multitude of instruments and actors involved exacerbate the well-known attribution problems between inputs/actions and outputs/systemic changes.

- In the same vein, the time span between the initiation of change through the various measures of the respective 'policy-mix' and the effects (especially the ones on the 'system/regime' level) can be very long and beyond the scope of current monitoring and evaluation techniques.

Still, in our view, ex-post evaluation of mission-oriented policies has the potential to 'trace 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. It would be of limited value as a tool for investment decisions, though.

CONCLUSIONS AND FUTURE PERSPECTIVES

MISSION-ORIENTED ASSESSMENT AND EVALUATION IN PRACTICE

A first screening of current practices in assessing and evaluating mission-oriented RD&I programmes has shown that very few such exercises actually have been conducted so far. However, first steps have been made in countries with explicitly mission-oriented programmes, in particular Austria, Denmark, Finland, Germany, Sweden and the UK.

In Austria, for instance, the high level of aspiration of RD&I funding programmes in terms of contributing to mission goals has led to the formulation of demanding requirements for their assessment and subsequent evaluation. Other countries like Denmark have started to explore the requirements for future assessments and evaluation in line with their strategy (DMSIHE 2012; DCSR 2013)

However, there is currently no systematic overview of „good practices" of assessing and evaluating mission-oriented RD&I programmes available yet. As new approaches are being tested, these should be carefully monitored. There is definitively a need to broaden the information base on such assessments and evaluations. In view of the methodological challenges associated to assessing and evaluating mission-oriented programmes, there is also a need for exploring new directions of policy research. New approaches to impact assessment need to be developed and tested, including new types of system modelling that allow capturing the complexity of impact pathways and scenarios in systems of innovation, production and consumption.

At the same time, the inherent limits to impact assessment need to be recognized and accepted. Evaluations as well as impact assessments should also build on a broader range of dimensions of analysis, in line with the range of mission-oriented goals. Economic impacts are just dimension to consider, next to social, environmental and other dimensions. Finally, in view of the long-term impacts to be considered, iterative processes of learning and adjustment need to be put in place, drawing on the insights from impact assessments and evaluations.

THE WAY FORWARD – A PESCA APPROACH FOR NEW MISSION-ORIENTED PROGRAMS

In this contribution we have argued for a new approach in impact assessment when dealing with new mission-oriented policies. We have stressed that the far larger complexity of these types of policies raise the stakes for impact assessment considerably, but we think that they can be tackled. For this purpose we have proposed a framework which puts much focus on

- Ex-ante impact assessment, based on scenario approaches and potential impact pathways, with a strong component of 'Social Cost-Benefit-Analysis';

- The establishment of sound relations between instruments and mission-goals upfront;

- An iterative-formative assessment process, which allows for the adjustment of objectives and instruments over longer periods of time to take account of
    - new technological possibilities;
    - better understanding of technological and economic potentials and limitations;
    - changing perceptions and needs of society;
- Ex-post evaluation in this frame would serve rather as a tool for historical 'critical path analysis' to identify the key drivers which were responsible for the success/failure of a specific policy than as one by which to rank investment priorities

A frame of reference which is broadened beyond RD&I, in order to cover also domain-specific policies (e.g. in transport, energy, health, …), will be essential if the scope of new missions is seriously interpreted as a transformative processes. We are confident that the approach we propose and that we would like to label the "PESCA (Prospective & Adaptive Societal Challenges Assessment) Approach", though demanding, would be a step forward in evidence-based mission-oriented policy making. It is a very much needed one, as current experiences with mission-oriented polices show (like the 'Energiewende' or Climate Change oriented policies amply demonstrate). And it's a worthy one given that the stakes are high.

REFERENCES

Collingridge, D. (1980): The Social Control of Technology, London: Pinter

DMSIHE (2012): Research 2020. Strategic Research Horizons, Danish Ministry of Science, Innovation and Higher Education, Copenhagen

DCSR (2013): Strategic Research. Principles and Instruments, Danish Council for Strategic Research, Copenhagen

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, Pages 1697-1792 (December 2012)

Gassler, H., Polt, W., Rammer, C. (2008): Priority Setting in Research & Technology Policy: Historical Development and Recent Trends. In: Nauewelaers, C. und Wintjes, R. (eds.): Innovation Policy in Europe. Measurement and Strategy, S. 203-224. Edward Elgar. Cheltenham.

Geyer, A., Rammer, C., Pointner, W., Polt, W., Hollenstein, H., Donzé, L., Arvanitis, S. (2000): Evaluierung des ITF-Programms FlexCIM, Wien

Renn, O., Klinke, A., van Asselt, M. (2011): Coping with Complexity, Uncertainty and Ambiguity in Risk Governance: A Synthesis

Soete, L., Arundel, A. (1993): An Integrated Approach to European Innovation and Technology Diffusion Policy. A Maastricht Memorandum, Brussels/Luxembourg

Tukker A., Chapter M., Vezzoli C., Sto E. & Munch Andersen M. (eds.): System Innovation for Sustainability. Perspective on Radical Changes to Sustainable Consumption and Production, Sheffield: Greenleaf

Weber, K.M., Johnston, R. (2008): Prospective strategic evaluation of national RTDI funding systems, Research Report, AIT/ACIIC, Vienna/Sydney

Weber M., Rohracher H. (2012): Legitimizing research, technology and innovation policies for transformative Change. Combining insights from innovation systems and multi-level perspective in a comprehensive 'failures' framework, Research Policy, Vol. 41, 1037 - 1047

AUTHORS

K. MATTHIAS WEBER
AIT Austrian Institute of Technology, Innovation Systems Department
E: matthias.weber@ait.ac.at

WOLFGANG POLT
Joanneum Research, POLICIES
E: wolfgang.polt@joanneum.at

ASSESSING THE BROAD SOCIETAL IMPACTS OF RESEARCH: THE CASE OF THE NCCR NORTH-SOUTH PROGRAMME[1]

CLAUDIA MICHEL, SIMON HEARN, GABRIELA WUELSER and THOMAS BREU

There is growing interest in the impact of academic research on society. If we define research impact as the "demonstrable contribution that excellent research makes to society and the economy", the concept encompasses a variety of contributions of research-related knowledge and skills that benefit people and the environment.[2] Prominent research networks such as the Research Councils UK, quoted above, are driving efforts to document the social and environmental benefits of research. Meanwhile, individual researchers from diverse disciplines are using their studies to address key issues – e.g. poverty, environmental degradation, or health burdens – and successfully helping solve societal problems. This trend towards emphasising the extra-academic benefits of research means that universities and researchers must contend with new expectations that go beyond those of scholarship and education. Some observers have begun using the term "third academic mission" to describe universities' efforts to engage with societal beneficiaries and achieve extra-academic returns (Göransson, Maharajh, and Schmoch, 2009).

In the following, we argue that research benefits society in a variety of ways, producing tangible returns over and above economic impacts, and that this has concrete implications for research evaluation. The Swiss National Centre of Competence in Research (NCCR) North-South provides a useful case example. It was a 12-year international research programme on sustainable development and global change that effected social, environmental, and economic returns around the world. Located in Bern, Switzerland, the programme's management centre developed and adapted its own self-assessment tools because the assessment instrument supplied by its main academic funder focused primarily on economic benefits, unnecessarily overlooking other contributions. After describing these assessment tools, we conclude our discussion by highlighting the potential and challenges of evaluating the diverse impacts of such research.

THE IMPACT OF RESEARCH ON SOCIETY: EMBRACING A BROADER PERSPECTIVE

The impact of research on society is often viewed primarily through an economic lens, focussing on wealth creation, productivity increases, profits, and strengthened global competitiveness. The EU's Framework Programme for Research and Innovation, for example, associates research and innovation with "successful commercialisation".[3] In Switzerland, the country's federal agency for the promotion of science-based innovation, or Commission for Technology and Innovation (CTI), emphasises entrepreneurship and the creation of start-up companies.[4] This stress on economic returns and enhanced cooperation between academia and industry dominates both on the national and the international level.

However, some observers argue that the impact of research must be considered more broadly. Claire Donovan, for example, describes research impact as belonging to a "social contract that exists between science and society…[and]…entails that research must address pressing social issues" (Donovan, 2011, p. 176). This new definition of research impact still includes economic returns, but it also accounts for social, cultural, and environmental dimensions. In this way, research impacts may be pursued on a variety of levels. Depending on their disciplinary background, individual researchers may measure the impact of their research in terms of environmental improvements, greater social cohesion, or reduced health burdens. Economic factors will continue to receive significant attention despite any new, broader conception of research impact. However, the well-being of societies in industrialised, emerging, and developing countries depends on more than just economic factors. Societal well-being also depends on factors such as social equity and the health of our natural environment.

1 This paper presents activities of the Swiss National Centre of Competence in Research (NCCR) North-South, funded by the Swiss National Science Foundation (SNSF), the Swiss Agency for Development and Cooperation (SDC), and the participating institutions.

2 Information on the British research councils' "Excellence with Impact" framework: http://www.rcuk.ac.uk/kei/Pages/home.aspx; retrieved on 4 March 2013

3 http://www.swisscore.org/SiteCollectionDocuments/Newsletter/syn_syn_1302.pdf; retrieved on 7 March 2013

4 Commission for Technology and Innovation CTI: http://www.kti.admin.ch; retrieved on 2 November 2013

BEYOND QUANTITATIVE APPROACHES TO IMPACT ASSESSMENT

Broadening our understanding of research impact has direct implications for research evaluation and the standards of quantitative and qualitative evaluation. Within academia, certain quantitative indicators have established themselves as the primary means for assessing research excellence. Publication counts, the impact-factor of publications, and competitive funds obtained, for example, have become popular proxies for research excellence (Donovan, 2007). These indicators may provide useful information about the resonance of particular research in the academic arena, but they say little about extra-academic returns. Other quantitative means for assessing extra-academic research returns have been developed, but they still have considerable flaws (Donovan, 2007). Many of these indicators and metrics have been adopted from the business world – such as level of industry funding, number of patents generated, or number of start-ups launched – and are commonly used for impact assessment by research funders and other key academic stakeholders. Rooted in economics, they fail to capture other important benefits that research may afford society. Finally, there have been efforts to create quantitative indicators that specifically assess societal returns, but these too are problematic. As Claire Donovan concludes: "The search for quantitative impact indicators has delivered an array of novel metrics that represent low-order impact, technometrics that privilege private over public interest, and sociometrics that rely on macro-level data with no credible link to the efforts of particular researchers" (Donovan, 2007, p. 591).

Qualitative methods of assessing extra-academic benefits are often appreciated for the greater flexibility they offer, enabling evaluators to account for various dimensions, including the public value of research. Experts in research assessment generally recommend combining quantitative and qualitative methods to evaluate research impacts, and they recommend peer review as a primary means of qualitative assessment. But peer review bears its own problems when used for broad impact assessment. Some observers point out that current practices of peer review are overwhelmingly based on discipline-specific value judgments. Disciplinary criteria of excellence are often poorly suited for assessing interdisciplinary research, or determining the societal relevance of research. According to Paul Nightingale and Alister Scott, "[t]he difference between the disciplinary emphasis of knowledge producers and the interdisciplinary needs of users is the most obvious relevance gap" (Nightingale and Scott, 2007, p. 545). They argue that research evaluation procedures such as peer review have contributed to expanding, rather than closing, the gap between the perceived quality of research and its actual relevance to society. Among other suggestions for improving research evaluation, they recommend furnishing reviewers with specific relevance criteria for use in the peer-review process, and offering reviewers guidance as to how to apply these criteria.

In sum, the prevailing methods used to assess the extra-academic impact of research focus too heavily on economic returns (neglecting social and environmental benefits), give undue weight to insufficient quantitative criteria, and/or are constrained by discipline-specific perspectives. Whether quantitative or qualitative, these assessment methods require further refinement in order to adequately capture the diverse ways that research may impact and benefit society.

THE NCCR NORTH-SOUTH: RESEARCH TO BENEFIT SOCIETY

Despite the continuing lack of adequate methods for assessing the broader societal impacts of research, individual researchers are still working hard to achieve such impacts – and learning how to assess them in the process. Below, we outline the experience of the NCCR North-South programme in developing a new approach for reporting and assessing the broader extra-academic impact of its research. The NCCR North-South was a transdisciplinary, international research programme based on partnerships between Swiss universities and other institutions in Africa, Asia, and Latin America (Hurni, Wiesmann, and with an international group of co-editors, 2010; Wiesmann U, Hurni H, and with an international group of co-editors, 2011). Comprising a network of around 1,200 researchers active in over 40 countries, the programme was dedicated to addressing challenges of global change and sustainable development. It received approximately 100 million Swiss francs in funding from 2001 to 2013, and enabled researchers to conduct advanced studies on topics such as livelihoods, institutions, conflicts, health, sanitation, economy, governance, and the sustainable use of natural resources.

As a research programme truly dedicated to improving human well-being and the environment, the NCCR North-South had a research mission that sought "to support societies in partner countries and institutions in their efforts to address syndromes in their regions and find means to mitigate them" (Hurni H, Breu T, Wiesmann U, and with contributions from the Board of Directors and the Management Centre of the NCCR North-South, 2013, p. 45). As a result, programme researchers were expected to strive for results that would benefit entire societies, not just marketable products to benefit the economy. Individual researchers conducted projects that aimed, among other things, to support more effective and efficient public services, more responsive policies, and improved understanding of global change.

Various structural conditions helped the research programme to achieve sustained, measurable impacts. The single most important factor was the programme's combination of funders: the Swiss National Science Foundation (SNSF) and the Swiss Agency for Development Cooperation (SDC) provided roughly matching funding to the programme.[5] These funding bodies – one oriented towards **academic excellence** (SNSF), the other towards **societal benefits** (SDC) – ensured that both academic rigor and extra-academic impacts were pursued over the programme's entire lifespan. In addition to this central supporting factor, four other key elements facilitated an enabling environment for impact creation. First, the programme's leaders had a shared understanding of the importance of societal impact, based on the research mission articulated above. Second, the programme's review panel, which evaluated it yearly and pro-

5 The participating Swiss universities helped fund the programme as well, but to a lesser extent.

vided feedback, was comprised not only of senior researchers, but also included representatives from international development work (albeit fewer in number). Third, the programme's funding scheme mandated establishment of a "knowledge and technology transfer" unit, whose aim was to channel relevant research products into the economy and society. Fourth, the programme's reporting scheme, provided by the SNSF, included sections for assessing academic quality and societal impact.

Nevertheless, the standardised SNSF reporting scheme could not account for the majority of the programme's societal benefits. The reporting scheme focussed on quantitative indicators designed to measure a research programme's economic benefit, such as the number of generated patents, licences, start-up companies, prototypes/demonstrators, processes/products, and CTI projects.[6] Aside from the number of processes and products it generated (98), the NCCR North-South programme performed badly on such metrics. The programme only produced one patent, five start-up companies, four prototypes or demonstrators, and no licences or CTI projects (Hurni H et al., 2013). However, the focus of the programme had been on knowledge transfer to policymakers and civil society actors. It aimed at generating research-related knowledge and skills that would benefit people and the environment. Outputs relevant to technology transfer (e.g. patents) or for-profit purposes (e.g. licences) were considered of minor importance vis-à-vis the programme's mission. In the end, while the review panel regularly expressed approval for the programme's societal impacts, their positive feedback was overshadowed by the programme's poor performance according to the standard reporting/assessment scheme of the SNSF.

RESEARCH IMPACT PLANNING AND SELF-ASSESSMENT: A NEW SWISS APPROACH

As a result of the mismatch between the SNSF's reporting/assessment scheme and the NCCR North-South's mission, the programme's management team developed and adapted instruments for reporting, planning, and assessing its impacts. It launched a series of NCCR North-South "reports on effectiveness". These publicly disseminated reports provided an overview of the programme's various impacts with respect to international development (Michel, Heim, Herweg, Zimmermann, and Breu, 2010), knowledge exchange between academic and non-academic actors (Heim, Michel, Salmi, and Breu, 2011; Michel et al., 2013), the career development of programme researchers (Heim et al., 2012), and maximising research impacts (Michel et al., 2013)

In addition to this series of impact reports, the management team introduced instruments to support programme researchers in maximising the effect of their engagement with societal beneficiaries. To aid the planning and monitoring of research impacts, the NCCR North-South adapted the RAPID Outcome Mapping Approach (ROMA). ROMA was developed by the Research and Policy in Development (RAPID) programme at the Overseas Development Institute (ODI). It is a novel approach for analysing and maximising research's impact on behalf of poverty reduction in developing countries (Young and Mendizabal, 2009). Its understanding of impact derives from Outcome Mapping, a methodology for planning, implementing, and evaluating development projects and programmes (Earl, Carden, and Smutylo, 2001). ROMA focuses on measuring the observable, behavioural outcomes that are necessary and sufficient for impact.

Outcome Mapping defines outcomes as "changes in the behaviour, relationships, activities or actions of the people, groups and organisations with whom a programme works directly" (Earl et al., 2001, p. 1). It applies a systems-thinking perspective to position a research programme's outcomes in terms of its contribution to ongoing development processes. The non-academic partners with whom researchers work directly and with whom they anticipate opportunities for influence are essential in the systems in which they are engaging. Development is viewed as a product of people's relations with each other and their environment. ROMA helps researchers plan ways of positively influencing the behaviour of non-academic partners from the outset of their research project; it also supports continuous monitoring of results in all stages of research. Ultimately, the seven-step ROMA framework seeks to aid researchers in achieving lasting impacts by triggering changes in broader policy (Figure 1).

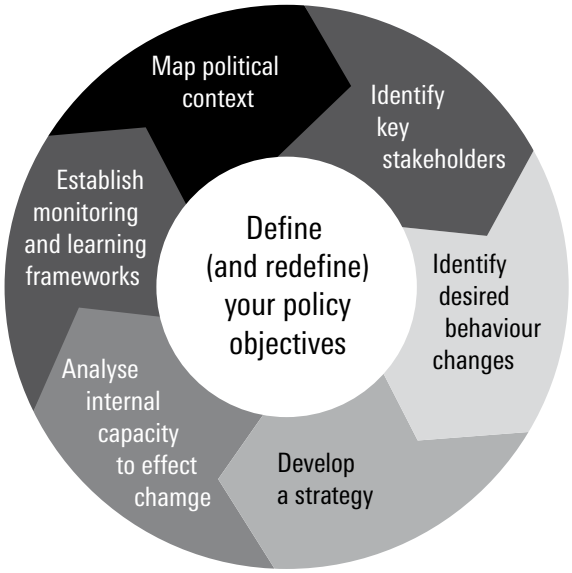


FIGURE 1 - The RAPID Outcome Mapping Approach (ROMA). (Source: Young and Mendizabal, 2009, slightly adapted. Reproduced with kind permission of the authors)

In contrast to the SNSF reporting/assessment scheme, the ROMA approach was well suited to the NCCR North-South's impact goals. Programme researchers found the ROMA tools and instruments useful. In 2012, for example, a group of NCCR North-South postdoctoral researchers jointly reflected on the benefits and limits of applying the ROMA approach within their research projects. Using the approach, they identified desired research effects in a variety of fields such as health policy in Tanzania and Chad, urban planning in Bolivia, and natural resource management in Tajikistan and Pakistan. ROMA enabled them to rapid-

6 CTI projects are projects financed by the Commission of Technology and Innovation (http://www.kti.admin.ch; website visited on 29 November 2013

ly identify their objectives in terms of influencing policy and to present key findings to policymakers in a comprehensible way. It helped them structure their ideas, emphasise the role of stakeholders, and focus on outcomes. Many felt that the approach supported them in better translating research into action for the benefit of their societal stakeholders (Michel et al., 2013).

NCCR North-South researchers generally appreciated the ROMA approach, but certain steps proved challenging. The most difficult step for researchers was that of identifying desired changes in behaviour on the part of key societal stakeholders. Researchers found it hard to specify how people's behaviours, relationships, activities, or actions should be transformed based on successful researcher–stakeholder interaction. Researchers were asked to clearly delineate how changes could be evaluated using short-term to long-term qualitative indicators. Despite the fact that programme researchers were trained in collaborating with non-academic stakeholders and in assuming different roles vis-à-vis research users (Pohl et al., 2010), they still found this task difficult. What clearly emerged from the group's collective reflection was the need for even stronger engagement between research evaluation specialists and researchers in order to develop coherent, effective mechanisms for self-evaluating the impacts of the NCCR North-South and similar research programmes.

CONCLUSION

Researchers, programme designs, and donor strategies are increasingly drawing attention to the potential impact of research on society. In many cases, however, economic benefits receive the most focus when assessing research impacts, and social and environmental benefits are overlooked. Indeed, the prevailing methods of research evaluation are of limited value for assessing the diverse societal impacts that research may have, especially research on sustainable development. Some of the popular quantitative metrics that are used to evaluate research have been adopted from the business world and macroeconomics, and favour private over public interests. Qualitative approaches such as peer review generally offer more flexibility in research evaluation, enabling adaptation of criteria to specific contexts and complex issues. Yet the current standard procedures of qualitative research assessment are overwhelmingly based on discipline-specific value judgements. In order to adequately evaluate the impacts of interdisciplinary research, for example, qualitative assessment procedures like peer review require further refinement. As regards quantitative approaches to research evaluation, the NCCR North-South programme provides an instructive example of the limitations of commonly applied models, such as that used by its primary academic funder the Swiss National Science Foundation (SNSF). The SNSF's reporting/assessment scheme focussed on economic returns such as generated patents, licences, or start-up companies.

Several key factors enabled the programme to strive for and accomplish societal impacts: long-term co-funding by an academic and, especially, an extra-academic development-focused funding body; a mixed review panel comprising academic and non-academic members; a programme design with a clear societal mission at its core; and participating researchers committed to engagement with non-academic stakeholders/societal beneficiaries. In the process of pursuing its societal mission through research, the programme's management team developed a new reporting procedure to aid impact assessment. Further, it adopted the ROMA planning, implementation, and monitoring tools to maximise the impact of the programme. Individual researchers appreciated these tools, but found that additional research evaluation expertise was necessary to realise their full potential.

Several other relevant lessons may be drawn from NCCR North-South programme. The demand that academic research beneficially impacts broader society requires a fundamental shift in research orientation. This shift in focus from academic achievements to extra-academic impacts cannot be delegated to lone researchers and cannot be treated as a supplementary, voluntary task. It requires establishment of additional tools and policies at the highest levels of research institutes and universities, as well as on the level of Science, Technology and Innovation (STI) policies. At present, research institutes and universities generally do not sufficiently incentivise or reward research that strives for extra-academic impacts. In Switzerland, the STI policy framework does still not embrace a broad understanding of research impact that accounts for societal and environmental returns, in addition to economic ones. Indeed, greater political will and institutional resolve are needed in order to bridge the gap between academic research and society.

REFERENCES

Donovan, C. (2007). Introduction: Future pathways for science policy and research assessment: Metrics vs peer review, quality vs impact. Science and Public Policy, 34(8), 538-542.

Donovan, C. (2011). State of the art in assessing research impact: introduction to a special issue. Research Evaluation, 20(3), 175-179.

Earl, S., Carden, F., and **Smutylo, T.** (2001). Outcome mapping. Building learning and reflection into development programs. Ottawa: International Development Research Centre.

Göransson, B., Maharajh, R., and **Schmoch, U.** (2009). New challenges for universities beyond education and research. Science and Public Policy, 36(2), 83-84.

Heim, E. M., Engelage, S., Zimmermann, A. B., Herweg, K., Michel, C., and **Breu, T.** (2012). Tracking Alumni Career Paths: Third NCCR North-South Report on Effectiveness, NCCR North-South Dialogue. Bern, Switzerland: NCCR North-South.

Heim, E. M., Michel, C., Salmi, A., and **Breu, T.** (2011). Does it Work in Practice? Fostering Knowledge Exchange for Sustainable Development. Second NCCR North-South Report on Effectiveness, NCCR North-South Dialogue. Bern: NCCR North-South.

Hurni H, Breu T, Wiesmann U, and with contributions from the Board of Directors and the Management Centre of the NCCR North-South (2013). NCCR North-South Final Report, 1.7.2001 - 30.6.2013. Bern, Switzerland: NCCR North-South.

Hurni, H., Wiesmann, U., and **with an international group of co-editors** (Eds.) (2010). Global change and sustainable development: A syn-

thesis of regional experiences from research partnerships (Vol. 5). Bern, Switzerland: Geographica Bernensia.

Michel, C., Heim, E. M., Herweg, K., Zimmermann, A. B., and **Breu, T.** (2010). Making research relevant to policy-makers, development actors, and local communities. NCCR North-South report on effectiveness of research for development, NCCR North-South Dialogue. Bern, Switzerland: NCCR North-South.

Michel, C., Heim, E. M., Zimmermann, A. B., Herweg, K., and **Breu, T.** (Eds) (2013). Maximising the Impact of Research: The NCCR North-South Approach. Fourth NCCR North-South Report on Effectiveness, NCCR North-South Dialogue. Bern, Switzerland: NCCR North-South.

Michel, C., Paulsson, M., Heim, E. M., Zimmermann, A., Herweg, K., and **Breu, T.** (2013). Promoting Research-User Collaboration: An Assessment of Funding Schemes. Fifth NCCR North-South Report on Effectiveness, NCCR North-South Dialogue. Bern, Switzerland: NCCR North-South.

Nightingale, P., and **Scott, A.** (2007). Peer review and the relevance gap: Ten suggestions for policy-makers. Science and Public Policy, 34(8), 543-553

Pohl, C., Rist, S., Zimmermann, A., Fry, P., Gurung, G. S., Schneider, F. and **Wiesmann, U.** (2010). Researchers' roles in knowledge co-production: Experience from sustainability research in Kenya, Switzerland, Bolivia and Nepal. Science and Public Policy, 37(4), 267–281.

Wiesmann U, Hurni H, and **with an international group of co-editors** (Eds.) (2011). Research for Sustainable Development: Foundations, Experiences, and Perspectives (Vol. 6). Bern: University of Bern: Geographica Bernensia.

Young, J., and **Mendizabal, E.** (2009). Helping researchers become policy entrepreneurs. How to develop engagement strategies for evidence-based policy-making (pp. 4). London: ODI.

AUTHORS

CLAUDIA MICHEL
Centre for Development and Environment (CDE), University of Bern
E: claudia.michel@cde.unibe.ch

SIMON HEARN
Overseas Development Institute (ODI)
E: S.Hearn@odi.org.uk

GABRIELA WUELSER
Institute for Environmental Decisions, ETH Zurich
E: gabriela.wuelser@env.ethz.ch

THOMAS BREU
Centre for Development and Environment (CDE), University of Bern
E: thomas.breu@cde.unibe.ch

CAN POLICY DESIGN HELP ORGANIZATIONS IMPROVE THEIR NETWORKING CAPABILITIES? AN EMPIRICAL ANALYSIS ON A REGIONAL POLICY

FEDERICA ROSSI, ANNALISA CALOFFI, MARGHERITA RUSSO

In parallel with the interest in networks of innovation on the part of the academic literature, policymakers are increasingly recognizing the important systemic nature of innovation processes, involving many agents often engaged in networks of relationships (OECD, 1997; Mytelka and Smith, 2002; European Commission, 2003; Nauwelaers and Wintjes, 2008), and they are increasingly supporting the creation of networks among firms and other types of organizations. Examples are the EU Framework Programmes (Breschi and Malerba, 2009; Tindemans, 2009) as well as the many national and regional policies launched in the past decade or so (Branstetter and Sakakibara, 2002; Caloghirou et al, 2004; Russo and Rossi, 2009; Bellandi and Caloffi, 2010; Cunningham and Ramlogan, 2012).

Policies for innovation networks usually aim to support joint R&D, technological development or technology transfer projects or even, sometimes, networking per se (with a view to create a "critical mass" of experts or users in a certain technology). At the same time, these policy interventions may also help the participants improve their ability to perform collaborative innovation, by allowing them to gain experience in working with external partners on a specific activity. Such behavioural outcomes, while not generally considered the main objective of these policies, have the potential to generate long-lasting beneficial changes in the participants' competences and abilities (Gök and Edler, 2012).

An important question for policy design is what kind of networks should be supported, if the objective of the policy is not just to fund "successful" innovation projects, but also to increase the participants' ability to engage in collaborative innovation. Should policies simply provide funding to innovation networks on the basis of an assessment of the project they intend to realize, or should they promote the setup of networks with specific features, in order to increase the agents' innovative potential through networking?

POLICY CONSTRAINTS AND COLLABORATIVE INNOVATION

In order to investigate this question, we use a rich dataset on all the organizations participating in a set of regional policy programmes implemented in Tuscany (Italy) between 2002 and 2008. Some of these programmes imposed certain compulsory requirements on the composition of the innovation networks to be funded (in terms of the size of the partnerships and of the types of organizations that they should include), while other programmes left the participants free to organize their partnerships according to their needs. In comparing the two different groups of programmes, we analyse the effects of such constraints upon the participants' ability to engage in subsequent collaborative innovation.

We can expect constraints to have both negative and positive effects on learning. Constraints impose an additional layer of rules that may be misaligned with the participants' actual needs. If such rules are irrelevant, they may increase the transaction costs in the process of network formation. But such rules may even be detrimental, if they hamper the networks' innovative performance and learning processes. For example, networks may be required to involve a type of organization that is not necessary for the success of the project, and which may even have an adverse impact on it, or a large number of partners that create congestion and hamper communication, thus reducing performance.

Conversely, constraints may be instrumental in enhancing the participants' ability to engage in further collaborative innovation. By participating in relatively large and heterogeneous networks, organizations may become acquainted with a variety of partners (who can provide them with further networking opportunities) and they may gain experience in interacting with agents characterized by different competencies, cognitive frames and modes of operation. We analyse whether policy constraints have had an impact on the participants' collaborative innovation capabilities by focusing precisely on these aspects – the ability to form new networks and the ability to form more heterogeneous and larger networks – as evidenced by the participants' involvement in subsequent policy-supported innovation networks.

THE REGIONAL POLICY PROGRAMMES

Tuscany's regional government has been one of the most active promoters of innovation network policies in Italy. In the programming period 2000-2006 it promoted nine consecutive waves of four policy programmes, supported by European Regional Development funds (ERDF), funding innovative projects carried out by networks of organizations. Overall, the nine waves were assigned almost € 37 million, representing around 40% of the total funds spent on innovation policies in that programming period. 168 projects were funded, and carried out in the years 2002-2008.

In our analysis we shall consider only the funded projects[1]. While the overall number of participations amounted to 2,006, many organizations (348) had taken part in more than one project, so that the different organizations involved in the nine waves were 1,127. Table 1 shows the numbers and shares of participations and organizations involved in the programmes, classified into nine categories according to their nature: firms, business service providers (generally private companies); private research companies; local (business) associations; universities (and other public research providers); innovation centres (generally publicly funded or funded via public-private partnerships); chambers of commerce; local governments; and other public bodies. The largest share of participating enterprises were manufacturing companies (68%): of these, 21.8% were micro and small firms in the traditional industries of the region (marble production and carving, textiles, mechanics, jewellery). Micro firms in the service sector were an active group, with 1.8 projects each on average. Not all types of organizations were permitted to receive funding: large companies and organizations based outside the region could enter the projects only with their own resources.

TYPE OF ORGANIZATION	PARTICIPATIONS		PARTICIPATING ORGANIZATIONS		TOTAL FUNDING		AVERAGE FUNDING PER ORGANIZATION
	n.	%	n.	%	€	%	€
FIRM	914	45.6	680	60.3	13,348,181	36.3	19,630
UNIVERSITY	261	13.0	93	8.3	7,355,106	20.0	79,087
PRIVATE RESEARCH COMPANY	32	1.6	22	2.0	537,613	1.5	24,437
INNOVATION CENTRE	150	7.5	34	3.0	6,208,052	16.9	182,590
BUSINESS SERVICE PROVIDER	153	7.6	86	7.6	4,015,642	10.9	46,694
LOCAL GOVERNMENT	176	8.8	77	6.8	691,654	1.9	8,983
LOCAL ASSOCIATION	209	10.4	85	7.5	3,016,694	8.2	35,491
CHAMBER OF COMMERCE	49	2.4	11	1.0	802,151	2.2	72,923
OTHER PUBLIC BODY	62	3.1	39	3.5	815,448	2.2	20,909
TOTAL	2,006	100.0	1,127	100.0	36,790,543	100.0	32,645

TABLE 1 - Participants, agents and funding by type of organization

The various programmes addressed a set of technology/industry targets. A large share of funds was committed to widening the adoption of ICT and multimedia in traditional industries and SMEs (48.2%). Projects in opto-electronics, an important competence network in the region, received 16.4% of funds, while projects in mechanics received 7.5%. The remaining areas included organic chemistry (5%), biotech (4%), and others (new materials, nanotechnologies and combinations of the previously mentioned technologies).

The set of policy programmes can be divided into two major periods. The first, which included the majority of waves and participants, ran from 2002 to 2005, and absorbed 45% of the resources for the network policies. It included three programmes, divided into six waves. The second period started in 2006, and ended with the last intervention implemented in 2008. It included two programmes, divided into three waves. Out of the six waves launched in the first period (2002-2005), five were characterized by the imposition of several constraints which were not present in any of the waves in the second period (2006-2008). Table 2 shows the types of constraint characterizing the different waves: whether the programme demanded a certain composition of the partnership in terms of types of organizations involved (henceforth "minimum heterogeneity constraint"), and whether the programme demanded a minimum number of partners, greater than that implied by the heterogeneity constraint (henceforth "minimum size constraint").

1 See Russo and Rossi (2009) for a comparative analysis of funded and not funded project applications submitted to the RPIA_ITT programme.

WAVE	POLICY PROGRAMME	TYPE OF CONSTRAINTS:				
		MINIMUM SIZE OF THE PARTNERSHIP	MINIMUM NUMBER OF:			
			SMES	RESEARCH ORG.	INNOVATION CENTRES	LOCAL GOVERNMENTS
2002_ITT	RPIA 2002	6	4	1		
2002_171	SPD line 171		4		1	
2002_172	SPD line 172		4		1	
2004_171	SPD line 171	4	1			
2004_171E	SPD line 171					
2005_171	SPD line 171	10	5	1		1
2006_VIN	RPIA 2006					
2007_171	SPD line 171					
2008_171	SPD line 171					

TABLE 2 - Types of constraints in the different waves

THE EFFECTS OF POLICY CONSTRAINTS

The following figure 1 shows the heterogeneity and size of networks in a scatter diagram that distinguishes between programmes with and without constraints. To compute the heterogeneity of each network we have used the reciprocal of the Herfindahl index on the shares of participants belonging to the different categories of agents, while the network size is defined in terms of number of participants. The average size and heterogeneity of networks were greater when constraints were present.

Obviously, these comparisons do not tell us what the effects of constraints are: the features of networks in each programme may be influenced by many other elements (the amount of funds available, the technology area that the policy was designed to implement, the duration of the programme, and so on). Moreover, this approach does not allow us to distinguish between the effects of each constraint. While the constraints were strongly overlapping, they had different intensities in different programmes, and they were only loosely related: the programmes that imposed a highest minimum size were not necessarily those that imposed the highest heterogeneity, and vice versa programmes with low minimum size requirements may have had more strict heterogeneity constraints.

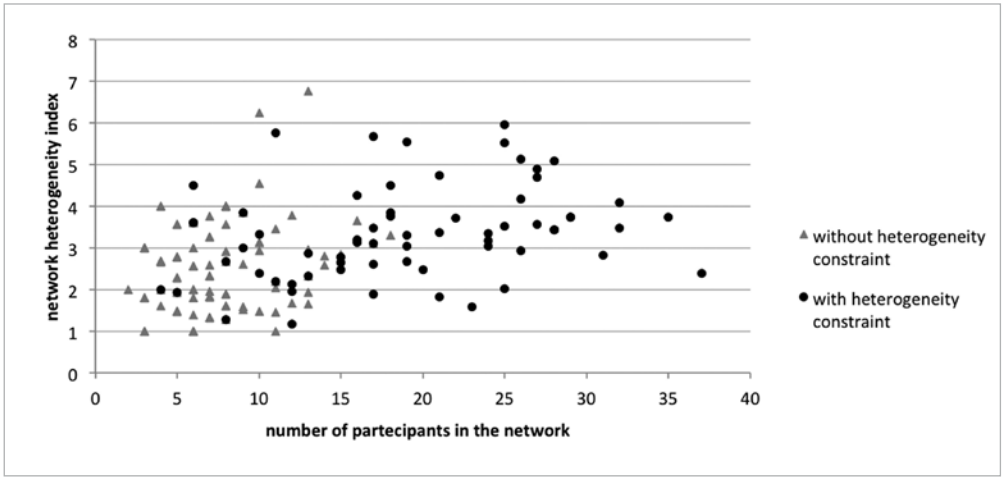


FIGURE 1 - Size and heterogeneity of project networks, grouped according to presence or absence of constraints

In programmes without constraints, network size was generally smaller and, although network heterogeneity was on average lower, its variability was greater.

In what follows, we try to explore the effects of policy constraints on the behaviour of each organization rather than on the behaviour of the networks of organizations. For each organization, we average the hete-

rogeneity indexes and the size of all the networks in which it took part, in either the first or the second period. The impact of constraints is also measured at the level of each organization: we compute the minimum heterogeneity requirements and the minimum size requirements of all the networks an organization participated in (where present), and we average these across all such networks.

First, we consider the 856 organizations that participated in programmes in the first period, and we assess whether policy constraints influenced the likelihood to participate also in the second period (Model 1). The dependent variable (*T_20068*) takes value 1 if the organization has participated in at least one project in the second period, and zero otherwise. Our hypothesis is that the policy constraints are likely to impact the actual heterogeneity and size of the networks the organization participated in during the first period, and these in turn are likely to affect the probability of its participation in the second period. To test this hypothesis we run a two-step instrumental variables probit regression (ivprobit) where the average heterogeneity and average size of networks in the first period (*avghet_20025* and *avgsize_20025*) are instrumented by the average minimum heterogeneity (*avgminhet_20025*) and the average minimum size (*avgminsize_20025*) constraints of the projects the organization participated in. We also include some variables capturing the organization's pre-existing capabilities for collaborative innovation (the number of projects the organization participated in during the first period, *Nprojects_20025*, and the average funding per project the organization was able to procure, *avgfunding_20025*), and we control for the organization's type and technological specialization (share of projects in each technology area).

Table 3 (page 20) reports the signs of significant coefficients found for Model 1. The first-stage regressions on the variables avghet_20025 and avgsize_20025 show that policy constraints significantly influence the heterogeneity and size of the networks each organization participates in: the variable avgminsize_20025 has a positive and significant coefficient in both cases, indicating that participating in networks that, on average, have higher minimum size requirements leads organizations to form larger and more heterogeneous networks. Instead, the variable avgminhet_20025 has a significant but negative coefficient in both cases, indicating that participating in networks that, on average, have higher minimum heterogeneity requirements leads organizations to form smaller and less heterogeneous networks.

Firms are involved in less heterogeneous networks, while several technological areas are positively associated with heterogeneity. Organizations that capture larger funds, on average, are involved in larger networks, and so are various types of organizations and several technological areas.

Concerning the main equation, neither greater heterogeneity nor greater size are associated with greater likelihood to participate in projects in the second period. Subsequent participation is more likely if organizations have obtained more funds and have participated in more projects in the first period, variables that can indicate the presence of stronger pre-existing collaborative innovation capabilities. The participation in a large number of projects in the first period may have further increased their collaborative innovation capabilities by providing them with more contacts and greater reputation as successful collaboration partners.

The result that participation in programmes with tighter minimum heterogeneity constraints had a negative effect on the heterogeneity and size of the networks presented, can appear counterintuitive. A possible explanation is that the specification of a more stringent constraint may have discouraged participants from including in their networks organizations that were different from the types recommended by the policymaker; that is, when confronted with very specific requirements, participants followed the guidelines for network composition quite closely and did not involve other types of organizations. This, paradoxically, led them to form networks that were less heterogeneous and smaller than those they may have formed had the constraint been looser (or absent). This interpretation is consistent with the observation that in programmes where heterogeneity constraints were present there was less variability in the project networks' heterogeneity indexes (see Figure 1) leading us to suggest that one of the effects of the heterogeneity constraints might have been to reduce the variety in the compositions of the different networks.

Secondly, we consider the set of 476 organizations that participated in the second period (2006-2008) and we examine whether having participated in projects in the first period that mandated constraints influenced three different characteristics of an organization's networks in the second period: the number of projects, Nprojects_20068 (Model 2), the average heterogeneity of project networks, avghet_20068 (Model 3), and their average size, avgsize_20068 (Model 4)[2]. Due to some missing data, the models are run on 460 observations.

The signs of significant coefficients found for Models 2, 3 and 4 are reported in Table 4. Model 2 suggests that having participated in projects with minimum heterogeneity and size constraints (*avgminhet_20025* and *avgminsize_20025*) did not influence the number of projects that the organization participated in during the second period. Rather, pre-existing collaborative innovation capabilities (*Nprojects_20025*) significantly and positively influenced the number of projects an organization participates in: having participated in more projects in the first period increased not only the likelihood to participate in projects in the second period (as shown by Model 1) but also the number of projects an organization participated in.

Model 3 suggest that having participated in projects with minimum heterogeneity and/or minimum size constraints did not influence the average heterogeneity of projects in the second period. Having participated in a greater number of projects in the first period had a significantly negative effect on the heterogeneity of networks in the second period: more experienced organizations ended up joining or forming less heterogeneous networks. Organizations may not consider heterogeneity per se as a valuable attribute of project networks, but rather only value when it is indeed necessary for the project's success: this is supported by the fact that in the programmes implemented in the second period, where no constraints were imposed, the networks' composition was more variable (as shown in Figure 1).

Model 4 suggests that having participated in programmes with heterogeneity and size constraints in the first period did not influence the size

2 Because of the different types of dependent variables, Model 2 is estimated with a Poisson model while Models 3 and 4 use OLS.

of an organization's project networks in the second period. From the previous Figure 1, we know that project networks in the second period were on average much smaller than in the first period, indicating that the minimum size constraints had indeed been effective in forcing organizations to form larger partnerships than they would have formed otherwise.

	FIRST STAGE	FIRST STAGE	MAIN EQUATION
DEPENDENT VARIABLE	**avghet_20025**	**avgsize_20025**	**T_20068**
avghet_20025			
avgsize_20025			
avgminhet_20025	-	-	
avgminsize_20025	+	+	
avgfunding_20025		+	+
Nprojects_20025	-		+
Ent	-	+	
Opub		+	
LA		+	
SC		+	
LG		+	
Uni		+	
SP			
shareICT	+	+	-
shareOpto		+	-
shareMEch	+		
shareOrgChem		-	
shareBiotech	+		
shareNew	+		-
shareMulti	+		
shareNano		+	-
shareGeo	+	-	
shareOther		-	
constant	+	+	+
N. observations	856	856	856
Wald test of exogeneity: chi2(2) = 5.59 Prob > chi2 = 0.0612			

TABLE 3 - Signs of significant coefficients in Model 1

DEPENDENT VARIABLES:	MODEL 2 Nprojects_20068	MODEL 3 avghet_20068	MODEL 4 avgsize_20068
avgminhet_20025			
avgminsize_20025			
avgfunding_20068	+		
Nprojects_20025	+	-	
Ent			+
Opub			
LA	-	+	
SC		+	
LG	-	+	
Uni		+	
SP		+	
shareICT			
shareOpto	+	-	+
shareMEch			
shareOrgChem		-	-
shareBiotech	-	+	+
shareNew		+	
shareMulti		-	-
shareNano			
shareGeo			-
shareOther		-	-
constant	+	+	+
N. observations	460	460	460

TABLE 4 - Estimates for Models 2, 3 and 4

CONCLUSIONS

These findings suggest several remarks on the effectiveness of constraints in supporting learning processes on the part of organizations involved in policy initiatives. Some constraints – especially less restrictive ones like the imposition of a minimum size – encourage organizations to interact with a larger number of organizations than they would not otherwise have partnered with. Although this does not necessarily translate in greater participation to subsequent programmes or in the formation of more diverse or larger networks in the second period, these contacts may provide useful in other contexts and at future points in time. Instead, a more restrictive constraint like the minimum heterogeneity constraint appears to have had more controversial effects: having participated in programmes with tighter heterogeneity constraints led organizations to form less heterogeneous and smaller networks.

The argument here is that very specific constraints were interpreted by participants as being akin to "guidelines" that should be followed in order to bid successfully; hence, in programmes with strict heterogeneity constraints the compositions of projects networks were more similar to each other, and reflected quite closely the minimum composition required by the policymaker. Instead, looser (or even absent) heterogeneity constraints led participants to include the variety of organizations that they actually needed to realize their projects, producing greater variability in network composition and, on average, greater heterogeneity.

The problem with the *ex ante* definition of very specific heterogeneity constraints is that, while there is a general consensus on the benefits of heterogeneous networks, the nature of the agents that may best contribute to the partnership very much depends on the content of the project that the network intends to realize. Hence, the definition of specific constraints may force participants to include organizations whose involvement is not needed for the purposes of the project, creating unnecessary complications. Rigid rules may even discourage participants from experimenting with more varied approaches.

Together, these findings suggest that collaborative innovation capabilities are gained over a longer time span than the duration of individual programmes, and that the imposition of simple constraints on network structure is not sufficient to ensure the acquisition of such skills. This is particularly true for projects that have small scale and short duration such as the ones we have analysed. In order to support organizations' capabilities to engage in collaborative innovation, strategies other than the imposition of constraints on network structure may be more productive: for example, implementing outreach actions in order to encourage orga-

nizations to participate in more policy supported innovation networks, and designing additional measures in order to increase the organizations' learning opportunities (providing opportunities to meet other organizations, facilitating meetings between different types of organizations, providing opportunities for joint action, and so on).

ACKNOWLEDGEMENT

The authors wish to thank Tuscany's regional administration for their invaluable help in sourcing the data underpinning this investigation.

REFERENCES

Bellandi, M. and **Caloffi A**. (2010). An analysis of regional policies promoting networks for innovation, in European Planning Studies, 18(1): 67-82.

Branstetter, L. and **Sakakibara, M.** (2002). When do research consortia work well and why? Evidence from Japanese panel data. American Economic Review, 92: 143–159.

Breschi, S., Malerba, F. (2009) ERA and the role of networks, in Delanghe, H., Muldur, U., Soete, L. (Eds.) European Science and Technology Policy: Towards Integration or Fragmentation?, Cheltenham, Edward Elgar.

Caloghirou, Y., Vonortas, N.S., Ioannides, S. (2004). European collaboration in research and innovation. Business strategy and public policy. Cheltenham: Edward Elgar.

Cunningham, P., Ramlogan, R. (2012) The effects of innovation network policies, Manchester Institute of Innovation Research, Manchester Business School, University of Manchester.

European Commission (2003) A European initiative for growth. Investing in networks and knowledge for growth and employment, COM(2003)579, Bruxelles, EC.

Gok, A., Edler, J., 2012. The use of behavioural additionality evaluation in innovation policy making. Research Evaluation, 21 (4): 306-318.

Mytelka, L., Smith, K., 2002. Policy learning and innovation theory: an interactive and co-evolving process. Research Policy, 31: 1467-79

Nauwelaers, C., Wintjes, R. (2008) Innovation Policy in Europe, Cheltenham: Edward Elgar.

Organization for Economic Cooperation and Development (1997) National Innovation Systems, Paris, OECD.

Russo, M. and Rossi, F. (2009). Cooperation Partnerships and Innovation. A Complex System Perspective to the Design, Management and Evaluation of an EU Regional Innovation Policy Programme, Evaluation, 15 (1): 75–100.

Tindemans, P. (2009). Post-war research, education and innovation policy-making in Europe, in Delanghe, H., Muldur, U., Soete, L. (Eds.) European Science and Technology Policy: Towards Integration or Fragmentation?, Cheltenham, Edward Elgar.

AUTHORS

FEDERICA ROSSI (corresponding author)
School of Business, Economics and Informatics, Birkbeck, University of London
E: f.rossi@bbk.ac.uk

ANNALISA CALOFFI
Department of Economics and Business, Universita' di Padova
E: annalisa.caloffi@unipd.it

MARGHERITA RUSSO
Department of Economics, Universita' di Modena e Reggio Emilia
E: margherita.russo@unimore.it

EVALUATION OF R&D INSTITUTIONS IN UKRAINE – THE NEW APPROACH

TRACK: FRAMEWORKS AND APPROACHES FOR EVALUATING NEW STI POLICIES AND INSTRUMENTS

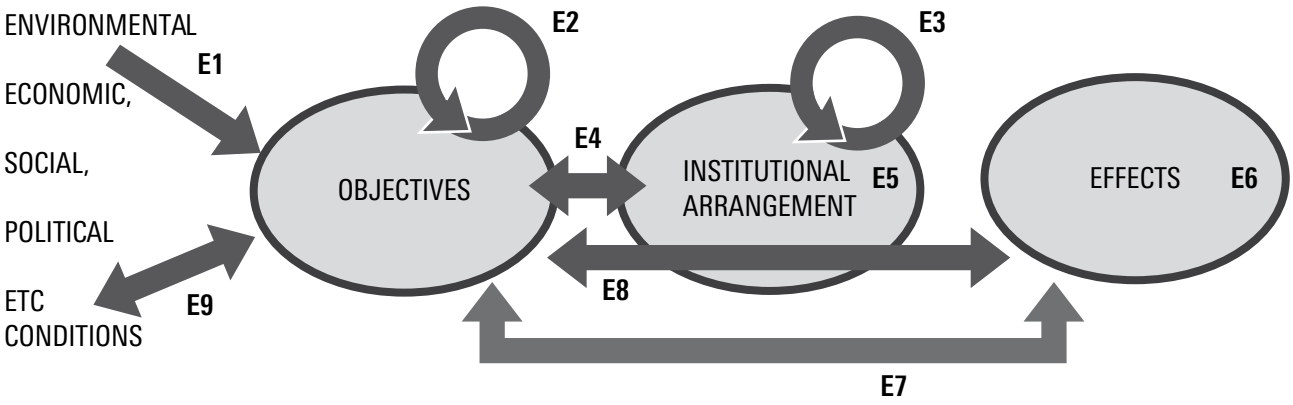
VITALII GRYGA, VICTOR RYBACHUK, OLHA KRASOVSKA

ABSTRACT

The main objective of the paper is to identify whether the current R&D evaluation system supports intensive S&T development of Ukraine. It was accomplished through employing a new methodology of evaluation of R&D institutes, which was developed under the Governmental Decree *"On Approval of the Concept of reforming the system of funding and management of scientific and technical activities"*. The new methodology allows making an assessment and is based on comparison of achievements in the previous period (10 years) and the dynamics of recent (4 years) trends in S&T and innovation activity of Ukrainian research institutions. The methodology is currently at the approbation stage, but about eighty R&D Ukrainian institutes have been already evaluated based on this new methodology. The results of this institutional evaluation are discussed in detail in this paper.

BACKGROUND

S&T policy evaluation is an important part of S&T policy implementation, which allows increasing its efficiency. The definition of "Evaluation", developed and used by OECD, is considered to be an *assessment, as systematic and objective as possible, of an ongoing or completed project, programme or policy, their design, implementation and results.* The aim of the evaluation is to determine the relevance and fulfilment of objectives, development efficiency, effectiveness, impact and sustainability. The evaluation should provide information that is credible and useful, enabling the incorporation of lessons learned into the decision making process of both recipients and donors (OECD DAC Glossary). According to this definition L.Bach (2011) proposed the following framework for S&T policy evaluation (fig. 1).



E1-E4 *denotes relevance of objectives and coherence between objectives and institutional arrangements*
E5 *relates to programme management, E6 – effects, outputs, outcomes*
E7-E8 *relates to match between objectives and relevant effect items*
E9 *represents ex-post relevance of objectives, given the results of other evaluations*

FIGURE 1 - Framework for S&T policy evaluation; *Source:* L.Bach (2011)

An evaluation of an institutional structure from the point of its coherence and conformity with the objectives of STI and R&D policies is one of the elements of the S&T policy evaluation framework. One of the key crucial problems of the field of research productivity evaluation as well as R&D institutional evaluation, is the objectification of evaluation results (Rybachuk, 2013).

Therefore, *the objective of the paper is* to identify whether the current R&D evaluation system supports intensive S&T development of Ukraine. To answer this question the following *specific aims* were established:

- to identify key differences and advantages of the current methodology of assessment of Ukrainian R&D institutes in comparison with the previous ones;
- to determine the main tendencies in evaluation of Ukrainian R&D institutes that give foundation for re-shaping the Ukrainian R&D landscape and in its turn for the improvement of S&T policy.

METHODOLOGY

There are different ways of institutional R&D assessments. Ranking is one of the approaches for evaluation of education and S&T activity. There are a lot of different rankings of universities in the world, i.e. Academic Ranking of World Universities, America's Best Colleges, Maclean's university rankings, CHE University Ranking (CHE-HochschulRanking), THE-QS World University Rankings, Ranking Australian Universities etc. At the same time rankings of R&D institutions are not widely used for national policy purposes. As a rule R&D units are evaluated inside the country, like the different research assessment exercises of scientific institutions employed in the United Kingdom or in the Russian Federation or Poland for example. Nevertheless different R&D institutions rankings in the global context exist, such as the Top American Research Universities, Ranking Web of World Research Centers, Web-sites rating of research institutes of the SB RAS etc. (The R&D report, 2012). In Ukraine however, evaluation systems of ranking research potential and research performance are used only for comparative assessment of higher education institutions. But there are no rankings of R&D institutions in Ukraine.

In Ukraine the evaluations of R&D institutions' performance are made in the following forms at state level:

a. statistical observations;
b. attestation of R&D institutions, supported by state;
c. information about results of public procurement on R&D and summary of its implementation monitoring.

Besides the above mentioned, there are additional ways of getting data for R&D evaluation purposes in Ukraine:

d. reports on performing of state targeted S&T programs, intersectoral and sectoral S&T programs, launched by the government. It reflects progress on the state program implementation according to its working plan, which consists of measures, corresponding financing, responsible executor and some performance indicators.
e. ex-ante evaluation of basic research proposals of project, that will be financed by the state budget. It is doing by Expert Council on evaluation of basic research projects' proposals. This Council is managed by National Academy of Sciences of Ukraine and consists mainly of academicians. Costs of the project, expected results, correspondences to priorities approved by the Government are the key issues have to be analyzed.

Also other types of evaluation are available, such as an assessment of economic efficiency of R&D costs and R&D results implementation in production as well as evaluation of innovation projects and technoparks performance. But in Ukraine annual statistical observation and collection of data on R&D could not be used as a tool for deep analysis of S&T policy due to several limitations. The reliability of statistical data is the first among them. These statistical observations are conducted through questionnaires, which are filled-in by R&D institutions themselves (self-assessment reports) and they tend to overestimate their performance (double counting of research papers, for example). That is why the Ukrainian government initiated new tools for evaluating R&D institutions. The state attestation of R&D institutions was implemented in 1998 (Decree of the Cabinet of Ministers of Ukraine, 1998), but it covered only those institutions which were included in the State register of research institutions, supported by the state[1]. The main weakness of this attestation was that it was based primarily on expert (qualitative) assessment and self-assessment questionnaires, which often led to overestimations and therefore such data could not be used as the foundation for R&D system reform initiated by the new President of Ukraine Viktor Yanukovich after the radical changes in Ukrainian government in 2009-2010. It stimulated measures for optimization of network of R&D institutions and improvement of scientometric evaluation of researchers. As a consequence the new methodology for R&D institutions evaluations was developed by the experts of the STEPS Center (Methodology of optimisation, 2011).

RESULTS

The new methodology is based on the methodological approaches to evaluate performance and effectiveness of S&T activities developed in the 1960s by G. Dobrov, a founder of scientometrics in Ukraine (G. Dobrov, 1966). The main goal of the new methodology is the comparison of achievements of the previous period (10 years) and the dynamics of modern (4 years) trends in S&T and innovation activity of research institutions. At the same time it takes into account former methods to make evaluation results comparable with the previous ones. The expert part of evaluation is to be provided by expert commissions established according to the basic types of S&T activities of research institutions. The methodology also allows identifying strengths and weaknesses of scientific institutions.

The main advantages of the new methodology are:

- the usage of quantitative as well as qualitative indicators (based on expert assessments);
- the usage of a formalized mechanism for limiting subjective evaluation ranking of scientific institutions: (a) priority to quantitative estimates of expert evaluation (through the use of differential weights - 0.6 and 0.4 respectively), and (b) nominalization of average rating. It allows to avoid the overestimation and influence of conformity that took place during the previous estimations;
- consideration of the differences between natural, engineering and socio-humanities sciences (through differential weights);
- combination of self-assessment and external assessment of research institutions;
- anonymous conducting of peer reviews and transparency of evaluation (through the on-line system "Expert").

1 R&D institutes included in this State register receive special preferences, first of all tax preferences (Decree of the Cabinet of Ministers of Ukraine, 2001)

The new methodology suggests assessing two main issues: *level of development* and *dynamics of institutional performance*. The system of indicators used in the new methodology is oriented first of all on the complex evaluation of S&T potential development. In its turn it should be used for increasing research quality, improvement of innovation activity as well as publication productivity, international recognition etc. (Rybachuk, 2012).

According to the Guide for evaluations of R&D institutions 42 indicators per institution were included in the evaluation process, and – in addition - 14 indicators were developed for the expert's assessment of R&D institutions.

15 indicators are used to assess the level of R&D institution's development, which reflect qualification of R&D staff, annual budget and research infrastructure, quality of research projects, patent activity and international recognition.

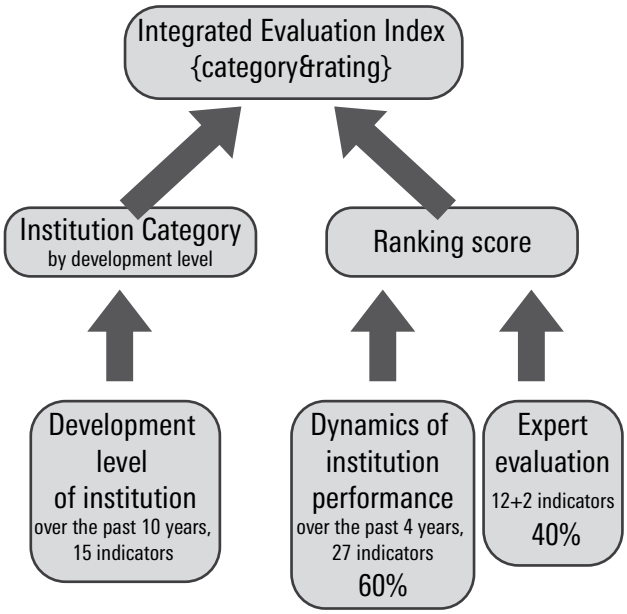


FIGURE 2 - The principal model of the evaluation process
Source: Developed by authors according to the Guide for R&D institutions evaluations, 2012

Criteria	Weight, %	Number of indicators	Main indicators
Qualification level of researchers	20	4	Number of highly qualified researchers (with scientific degree)
Budget and infrastructure	20	3	Total R&D funding (all sources) Number of unique objects of national heritage, number of centers of joint use of expensive equipment Number of teaching departments, laboratories, research and educational centers, which operate in research institution
Research projects profile	20	2	Number of state targeted R&D and S&T programs in which the research institution under scrutiny was involved during the last 10 years Number of international programs and projects in which the research institution under scrutiny was involved during the last 10 years
Productivity	20	2	Number of granted patents and sold licenses (in Ukraine and abroad) during recent 10 years
Prestige	20	4	Number of foreign scientometric database, in which research journals of the institutions are included Position in the "Ranking Web of World Research Centers" and "Ranking Web of World Universities"

TABLE 1 - Evaluation of the institutional development level; *Sources: Guide for R&D institutions evaluations, 2012*

As all mentioned indicators are in absolute numbers, larger institutes automatically have higher value in comparison with smaller ones. Indicators used for the assessment of an institution's development level allow taking into consideration differences between social sciences and humanities and natural sciences and engineering. Different weights are used to deal with them during the evaluation process as regards the following indicators:

	Weight (%) for natural sciences and engineering	Weight (%) for social sciences and humanities
R&D funding over the past 10 years, thousands UAH	10	15
Number of unique objects of national heritage, number of centers of joint use with expensive equipment	5	No
Number of books (monographs), encyclopedias, dictionaries and hand books (not including educational literature), published in Ukraine or abroad during recent 10 years	10	20
Number of granted patents and sold licenses (in Ukraine and abroad during recent 10 years	10	No

The dynamics of R&D institutions are analysed through evaluation of S&T potential and research productivity dynamics. For evaluating the S&T potential 13 indicators on human resources, financial, technical base and equipment, as well as innovations are used. Productivity's dynamics are analyzed by using 14 indicators on volume of research, publication activity, and international integration, including papers in international databases.

CRITERIA	NUMBER OF INDICATORS	MAIN INDICATORS
SUBINDEX I: DYNAMIC OF S&T POTENTIAL DEVELOPMENT		
Human potential	5	Share of highly qualified researchers (with scientific degree) to total number of research staff Share of young researchers till 35 years to total number of researchers Number of researchers' foreign trips for training, teaching and performing research for the period from 3 month to 2 years, per 100 of researchers
Financial, recourses and innovation potential	8	R&D costs per researcher Share of business sector financing in R&D financing Share of foreign funding in R&D costs Average annual volume of fixed assets of research activity Number of acting utility patents (both Ukrainian and foreign) Volume of dividends generated by companies and organizations, founded or co-founded by the institution
SUBINDEX II: RESEARCH OUTPUT INDICATORS		
Research volume	3	Share of financing for research in the framework of state intersectoral and sectoral programs, implementation of measures and tasks on priorities of S&T development in total budget financing The volume of applied research, developments and S&T services, conducted by own funding, per 100 researchers
Publications	2	Number of papers published in national and foreign research journals, which are included in internationally recognized databases per 1 researcher Total number of publications per full-time faculty
Innovation activity	4	Number of intellectual property protection documents, issued in Ukraine and abroad, per 100 researchers Number of license agreements on intellectual property use and know-how, per 100 researchers Number of innovation projects, in which the institution under scrutiny is involved, per 100 researchers Number of informational and analytical publications, drafts of state programs and legal documents prepared for central authorities by the institute
International integration	3	Number of projects in the frame of joint competitions with foreign partners, per researcher
Representation in the world information area	2	Number of pages of the institution website recovered from the search engine "Google" Number of papers published in national and foreign journals with high impact factor, per researcher

TABLE 2 - Evaluation of dynamics of institution performance; Source: Guide for R&D institutions evaluations, 2012

Dynamic evaluation is calculated as simple average of normalized (or standardized) values of average rates of dynamics of all indicators. For R&D institutions in natural sciences and engineering all 27 indicators are used. For social sciences and humanities the following indicators are not used:

- Number of valid utility patents granted in Ukraine and abroad
- Number of license agreements on intellectual property use and know-how per 100 researchers

The qualitative expert assessment is also included in the block of dynamic evaluation. It contains questions on conformity level of R&D institutes' research with state R&D priorities, the quality of research results, future prospects of the research activity in the institute etc.

Criteria	WEIGHTS (%)	NUMBER OF INDICATORS
Level of targeted orientation of research	16	2
Quality of research results	18	3
Prospects of R&D activity	56	7
Specific indicators	10	2
Total	100	12+2

TABLE 3 - The structure of expert evaluation; *Source: Guide for R&D institutions evaluations, 2012*

The ranking of the research institution is calculated considering quantitative and qualitative evaluations of its research capacities, dynamic of its development and level of research results as well as future prospects of the research results implementation into economy. Mathematically the ranking is calculated as a sum of dynamic and expert evaluation that are accordingly weighted on 0.6 and 0.4. The value lies in the interval between 100 and 500. The attestation procedure suggests formatting a specific matrix, which allows classifying research institutes into four categories.

Classification group	Classification evaluation (scores)	R&D institution ranking (scores)			
		100-150 Catching up	151-250 Moderate	251-350 Active	351-500 Leaders
I - II	2,61 - 5,00	Category C R&D institutions which have high S&T potential and are recognized in Ukraine; they participate in the S&T policy making process, but their developments' rate is slow.		Category A R&D institutions, which have high S&T potential and are recognized in Ukraine and globally. They take part in the S&T policy making process and influence sectoral S&T policy. They produce high quality results.	
III - IV	1,00 - 2,60	Category D Here is a place for R&D institutions with low S&T potential, and their development dynamic is low too. They are not prominent in the world research community. Such institutes conduct R&D in a very specific field and often on an ad hoc basis.		Category B S&T potential is lower than in Category A, but institutions show high rates of development and high efficiency. They are able to participate in policy making processes in a specific area. They are active in the world research community integration process and are among leaders in a narrow field, including implementation of research results.	

TABLE 4 - Matrix for R&D institution categories; *Source: Guide for R&D institutions evaluations, 2012*

In order to identify strengths and weaknesses of a research institute from the point of its development and productivity, ranking positions of mentioned 27 indicators should be calculated. The rating is identified by the order of decreasing the average rate of dynamic of institution performance. ĐĐResults of strengths and weaknesses analysis of the research institute activity should be taken into account during the process of identification of future prospects and preparing conclusions and applications on S&T system optimization as well as for increasing of research institution performance.

Comparative ranking evaluation allows decision-makers distinguishing losers and leaders among research institutions through rating them by ordinal scale. Such evaluation can be used by policy makers for the following purposes:

- to identify needs on adjustment, revision or additional expert evaluation of a research institution;
- adjustment of ranking values for more differentiation of attestation results;
- justification of conclusions in attestation results in cases of insufficient certainty, e.g. when a research institute does not agree with the results of the attestation.

It is suggested that the decision on a comparative ranking and its implementation should be made by a special expert commission established for implementing of the attestation results in practice. The characteristics of research institutions activity, received through attestation, and results of strengths and weaknesses analysis as well as factors, which determine them, will allow the expert commission to develop recommendations on actual directions and necessary measures for improvement of R&D activity of the institutions, including:

- level or volume of budget support, tax and others preferences;
- reorganization, changes in profile or termination of research institutions;
- other measurers on optimizing research institutions and the research system as a whole.

The State Agency on Science, Innovation and Informatization provides the Ministry of Science and Education information on the attestation and suggests own proposals on optimization of the network of research institutions in Ukraine based on the results of research institutes evaluation made by the expert commissions. These proposals are introduced to the Government where a final decision has to be made.

ÐOn the basis of the assessment made in accordance with the new methodology, a comparative ranking of selected groups of academic institutions could be done. It motivates research institutions to self objectification and implementation of effective measures to improve their activities, including justification of forms and volumes of state financial support. An evaluated research institute has to develop measures for further improvement of its activities that should be considered by the expert commission.

82 state research institutions, which are managed by different central authorities (and which do not belong to the national academies of sciences), were evaluated in 2012 according to the above-mentioned methodology. Information about each R&D institute was obtained on the following issues: efficiency, quality of R&D results, implementation of R&D results, S&T potential development, R&D orientation. As a result a ranking of these R&D institutions was established. The rankings of research institutions were in the range of values from 243 to 403 points. For 18 research institutions (23%) the rating was less than the average level (300 points).

According to the matrix (table 5), 32 research institutions were referred to the category A, 49 – to the category B, and only one was refered to the category D. There were no institutions in category C, which means that there are no R&D institutions in the sample which have a high S&T potential and are recognized in Ukraine but which with a slow development rate. This is quite reasonable, because an institution that has a strong potential and is recognized in the world will hardly have slow growth rates. When a large number of institutions will be evaluated by the new method, and this column will still be empty, the method may require a revision of its contents.

CLASSIFICATION GROUP	CLASSIFICATION EVALUATION (SCORES)	NUMBER OF INSTITUTIONS BY RANKING SCORE			
		100-150 Catching up	151-250 Moderate	251-350 Active	351-500 Leaders
I - II	2,61 - 5,00	Category C **No**		Category A **Total is 32**	
I	3,81 - 5,00	C3 **No**	C1 **No**	A3 **No**	A1 **2**
II	2,61 - 3,80	C4 **No**	C2 **No**	A4 **23**	A2 **7**
III - IV	1,00 - 2,60	Category D **Total is 1**		Category B **Total is 49**	
III	1,41 - 2,60	D3 **No**	D1 **No**	B3 **38**	B1 **8**
IV	1,00 - 1,40	D4 **No**	D2 **1**	B4 **3**	B2 **No**

TABLE 5 - The resulting matrix of 82 R&D institutions' distribution by categories
Source: Developed by authors from the materials meeting of the Board SASII, November 12, 2012

Another important issue is that usually expert evaluation has higher values than the evaluation of dynamics (see Fig. 3). At the same time it is more critical than the self-assessment of the institutes. Thus, 53 of 80 research institutes (or 66%) were evaluated by experts more strictly. Only for 11% of the institutes expert assessments were higher than self-estimations, and 22% expert assessments were equal with the self-estimations. It should be noted that state attestation is not just an exercise for mind or routine but it has practical applications. The results of the last evaluation have been already used in recommendations for the optimization of the state R&D institutions' network. Based on the recommendations a draft for a Governmental decree was developed which is currently in the stage of negotiation with relevant central bodies. At the same time the Decree of the Government on 28 November 2012 N° 983-p «On optimization of state research institutions network» contained a number of practical decisions of reorganization (e.g. merger) and even liquidation of some Ukrainian R&D institutes based on the results of the attestation.

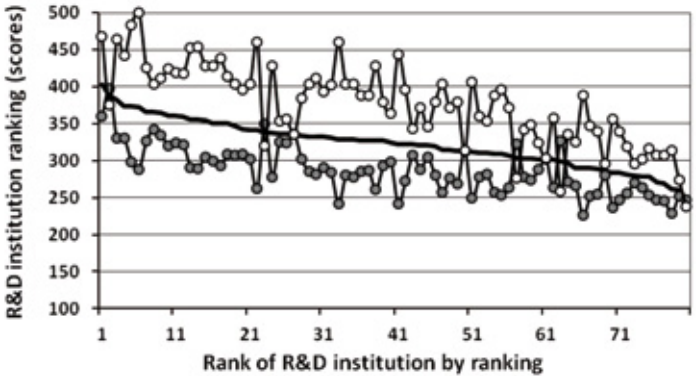


FIGURE 3 - Comparing the rankings of research institutions (curve without markers) with estimations of dynamics (grey marker) and experts evaluation (light marker)

CONCLUSIONS

The new methodology is still in the stage of approbation. Only 82 R&D institutes were evaluated under this methodology at the moment. The quality of the evaluation seems to have increased due to more orientation on expert evaluation and dynamic performance. In many cases new evaluations were lower than evaluations obtained through the old approach. The largest differences were between experts' opinion and self-evaluations, in particular in case of low ranking institutions.

Attestation of Ukrainian R&D institutes under the new methodology gives an opportunity to understand the main tendencies in the development of Ukrainian R&D institutes in order to re-shape the Ukrainian R&D landscape, to improve and to increase efficiency of S&T policy in Ukraine.

It is expected that the new methodology will stimulate research institutions to develop their potential and to increase their medium and long term performance. It also provides an opportunity to reveal unique strengths and weaknesses of R&D institutions and, thus, to justify measures for improving their activities.

The new methodology is not the final stop on the way of evaluating R&D institutions. Today there are more and more disputes on using impact factor and other bibliometric indicators in R&D evaluation in Ukraine, and nobody seems to know how to implement them by taking into account differences in engineering and humanities, for example, which databases (or journals) should be covered etc. These issues should be taken into consideration in the evaluation process in the future.

REFERENCES

OECD DAC Glossary. Available at: http://www.oecd.org/dac/dacglossaryofkeytermsandconcepts.htm

L. Bach, 2011, S&T Policy evaluation: focus on the impact evaluation/ Maastricht Session/Week 2 – Oct 2011. Available at: http://dimetic.dime-eu.org/dimetic_files/Maast2011LBach2.pdf

V. Rybachuk, 2013, Methodological issues of the S&T performance evaluation, Science and science of Science, Ð2, P. 46-52 (in Ukrainian).

The R&D report «Development of methodology for rating evaluation of scientific institutions for establishing and implementation of S&T expertise in Ukraine», 2012 / National Academy of Sciences of Ukraine, G.M. Dobrov Center for Science & Technology Potential and Science History Studies, Ðyiv, 101 p. (in Ukrainian).

Decree of the Cabinet of Ministers of Ukraine on 7 April 1998 N°469 "On approval of the Regulation on state attestation of R&D (S&T) institutions" / Accessible at http://zakon0.rada.gov.ua/laws/show/469-98-%D0%BF (in Ukrainian).

Decree of the Cabinet of Ministers of Ukraine on 23 April 2001 N°380 "On approval of the Regulation on the State registry of scientific institutions, which are supported by the state" / Accessible at http://zakon2.rada.gov.ua/laws/show/380-2001-%D0%BF (in Ukrainian).

Methodology of optimization (attestation) of budget scientific institutions, which are fully or partially financed by state budget / Approved by the decree of the Cabinet of Ministers of Ukraine on 31 October 2011 N° 1176. Accessible at http://zakon2.rada.gov.ua/laws/show/1176-2011-%D0%BF/ed20120403 (in Ukrainian).

G.M. Dobrov, 1966, The Science of Science. Introduction to the General Science of Science – Kiev: Naukova dumka (in Russian).

V. Rybachuk, 2012, Indicators of innovation activity in legal acts 2012 on national accreditation of research institutions in Ukraine // XVII International scientific and practical conference „The Problems and prospects of innovation development of the economy", Alushta (Crimea, Ukraine), 10–15 September 2012: Conference proceedings accessible at http://www.uiis.com.ua/files/Sbornik1%202012.pdf (in Ukrainian)

Guide for R&D institutions evaluations / Approved by Order of Ministry of education and science of Ukraine on 03 April 2012 N° 399. Accessible at http://zakon2.rada.gov.ua/laws/show/z0629-12 (in Ukrainian).

AUTHOR

OLHA KRASOVSKA
The State Fund for Fundamental Research of Ukraine
E: olha.krasovska@gmail.com

VITALII GRYGA, VICTOR RYBACHUK
STEPS Center, NAS of Ukraine
E: v.gryga@gmail.com
E: rybachuk.v.p@nas.gov.ua

PROMOTING FIRMS' INNOVATIVE BEHAVIOR IN RUSSIA: WHAT WEAKENS THE POWER OF STI POLICY?

YURI SIMACHEV, MIKHAIL KUZYK and VERA FEYGINA

INTRODUCTION

In recent decades, in the context of globalization and increasing international competition, national governments pay more and more attention to fostering innovation in order to compensate for market failures, system failures, and failures in the ability to absorb new knowledge. There is an extensive framework of innovation-fostering instruments: tax incentives, grants and subsidies, institutes for development, etc. However, the issues of efficiency and advisability of different innovation-supporting policies are crucial, especially in the situation of budget constraints: it is necessary to choose support instruments that would be efficient enough while minimally distorting the market environment.

Progress in learning new innovation support mechanisms and their application in Russian innovation policy has become apparent in recent years. Nevertheless, experience (both positive and negative) of application of earlier-introduced support policies is largely neglected when new policies are being designed. Moreover, expert discussions of improving public spending efficiency and the development of the Russian tax system have demonstrated that the debate of the pros and contras of diverse approaches to innovation support still brews in Russia, even at a very high level of generalization, e. g., when comparing tax and financial incentives.

For all its multidirectionality, innovation policy realized by the state still lacks one necessary attribute: regular independent progress evaluation. Although there is a general practice of macroeconomic evaluation of budget losses and short term benefits from the use of innovation support policies, there is no integrated system of analyzing the impact of different support mechanisms on innovation development at the micro level, such as change in companies' behavior, shifts in preferences of major stakeholders, developments in relations with research institutions, and increasing learning capacity.

In this context, the main objective of our paper is a micro-level study of how support mechanisms impact companies' innovation behavior. We also evaluate effectiveness of tax and fiscal policies employed by the state to support innovation and compare positive and negative effects of these policies' application.

BACKGROUND ON THE RUSSIAN STI POLICY

Over the last six years evident progress has been achieved in the development of Russia's STI policy. Nevertheless, no sustainable positive shift in the innovation sphere has occurred so far at the macro-level (Table 1). The majority of indicators have been fluctuating around very modest levels and demonstrated only slight increase over the period of 2006-2012. Moreover, the share of industry-financed R&D has even fallen. The only positive trend concerns the share of innovative goods and services in total sales. All in all, the percentage of organizations implementing technological innovations is still low; the role of business in financing research is very limited.

INDICATORS	2006	2007	2008	2009	2010	2011	2012
Gross domestic expenditure on R&D (GERD) as a percentage of GDP	1.07	1.12	1.04	1.25	1.13	1.09	1.12
Percentage of GERD financed by government	61.1	62.6	64.7	66.5	70.3	67.1	67,8
Percentage of GERD financed by industry	28.8	29.4	28.7	26.6	25.5	27.7	27,2
Enterprises engaged in technological innovation as a percentage of enterprises total	9.4	9.4	9.6	9.4	9.3	9.6	9.9
Expenditure on technological innovation as a percentage of total sales	1.4	1.2	1.4	1.9	1.5	1.5	1,8
Innovative goods and services as a percentage of total sales	5.5	5.5	5.1	4.6	4.9	6.1	7.8

TABLE 1 - Selected indicators of innovation activity in Russia, 2006-2012
Sources: HSE. (2012). Science. Innovations. Information-oriented Society: 2012. Higher School of Economics, Moscow; HSE. (2012). Science and Technology Indicators in the Russian Federation. Higher School of Economics, Moscow; HSE. (2012). Indicators of Innovation in the Russian Federation. Higher School of Economics, Moscow. www.gks.ru

Major achievements and constrains in the innovation system over the last six years are given in Table 2.

First of all, huge resources were available in the *pre-crisis period,* so that the state began to stimulate widely the demand for innovations. However, the policy toolkit was obviously biased towards direct support mechanisms. Thus Russian STI policy was focused on relatively big companies instead of supporting SMEs. In general, government expenditures on R&D were increasing significantly over the period, whereas actual innovativeness of the recipient projects were still out of view, so that business only imitated innovative behavior and at the same time showed signs of rent-seeking.

The crises period could be characterized by toughening budget constraints resulting in a lower level of investment in fixed assets and a sharp decline of innovation. At the peak of the crisis, budget expenditures were partly reoriented from innovative businesses to big strategic companies in order to maintain social stability. Various quotas were implemented for domestic market protection, and that, in turn, essentially worsened business environment. However, reevaluation of the role of innovations in Russia's national economy occurred during that period, so that a number of modern innovation support tools have been proposed.

One can observe the initiation of comprehensive instruments for the support of cooperation between different actors of the innovation system, fostering research at universities and the creation of networks taken place in the *post-crises period* (Dezhina, Simachev, 2012). State decision-making with regard to some crucial social and economic issues has been just recently implemented and in some areas the final decisions have not yet been elaborated (for example, tax policy, pension reform).

MACROLEVEL	MICROLEVEL
1. PRE-CRISIS PERIOD: 2007-2008	
• huge budget recourses • increasing public investment • growing budget allocations to innovation • adoption of long-term strategies, state targeted programs in science and technology • tax incentives for innovation • establishment of big venture funds	• stable conditions for business, a reduction of tax burden • risks of takeovers and discouragement of business expansion • mainly the imitative innovation model; low R&D spending • small number of truly innovative companies
Major constraints: broad application of 'rough' direct STI policy tools, strong distortions in the market environment	
2. CRISIS PERIOD: 2009-2010	
• dramatic budget curtailment; • countercyclical policy; • temporary protection policy, domestic demand promotion; • selective support of big companies; • establishment of state committees on modernization; • setting modernization priorities	• hard financial constraints for companies; • dramatic decline in the predictability of business environment conditions; • innovations are concentrated in the big businesses; • business is interested in costs reduction
Major constraints: 'confiscation' of potential advantages from innovatively-active companies due to the state policy's focus on social stability to the detriment of economic performance	
3. POST-CRISIS PERIOD: 2011 -2012	
• considerable budget constraints; welfare-oriented budget; • innovation is one of government policy's priorities; • significant alterations in regulation; • new innovation promotion instruments; • multiple 'experiments with no consequences' and learning from them	• uncertainty, low predictability of business environment; • multiple 'innovation signals' from the state; • businesses focus on completing their current projects; • imitation of innovation activity as a type of rent-seeking behavior; • increasing importance of launching new products
Major constraints: uncertainty of economic conditions; postponement of key economic decisions by the state; considerable slowdown in the institutional development of business environment	

TABLE 2 - Comparison of major achievements and problems of the STI policy in 2007-2012

Some experts also point out the negative influence of unstable business environment and low predictability of government economic policy on innovative growth. Thus, at present, the most relevant factors that hinder innovation activity of companies are, firstly, the unstable conditions for economic activity, which increase the risks and reduces the planning horizon; and secondly, the internal bureaucratization of the business processes inside companies, which makes them less open and receptive to innovations.

In this context, the empirical research has been designed in order to evaluate strengths and weaknesses of Russian STI policy and to understand better its influence on the companies' behavior.

DATA

The empirical base of the research was formed as a result of two surveys of Russian industrial enterprises, conducted in 2011 and 2012. In the first case the sample included 602 companies, in the second – 652 companies, representing mainly the manufacturing industry. The panel formed by the results of two surveys consists of 415 companies.

EMPIRICAL FINDINGS

1. COMPANIES' PRODUCTIVITY GROWTH: INNOVATION, IMITATION, OR OPTIMIZATION?

Innovation activity of an enterprise is often considered as a main factor in the increase of labor productivity (Bogetic, Olusi, 2013). However, this proposition has not been confirmed by our research sample: half of the companies who have increased their labor productivity have nothing to do with innovations. Innovative enterprises have been expanding production by means of increasing turnover and investment into renovation of production capacities, whereas a group of non-innovative companies have reached the same goal through staff reduction. We assume that excessive employment emerged due to the restrictions imposed on business during the recent crisis. Optimization of labor costs became possible for companies after the major phase of the crisis ended.

Therefore, we suppose that the persisting scarcity of innovation-active companies in the post-crisis period is the consequence of artificial constraints of business optimization. After these restrictions had been eased, companies received new opportunities to maintain their competitiveness, besides innovations. Choosing this path can be also the result of low cost and low importance of staff for innovation-passive and technologically backwards companies: there is no need in special skills in this case, and enterprises do not invest much in human capital development.

2. STATE SUPPORT OF COMPANY INNOVATION: IS THERE A NEW QUALITY?

The impact of innovative activity on companies' performance indicators has many channels, while it is not exactly productivity that is positively affected by innovations (OECD, 2011; Goldberg, 2011).

In general, among state support recipient companies, the situation almost always improved due to innovation, at least in some output indicators. This confirms the empirical observation that successful innovative companies receive public support more often compared to unsuccessful innovators (Hanel, 2003). However, this can also be attributed to the effects of selecting the best for the support.

We have found significant positive influence of state support on the increase in companies' exports and energy efficiency; however, no effect of state support on firms' productivity growth has been detected (Figure 1).

The research has shown, that the factor of state support per se is so far of no consequence for a firm's efficiency, especially for productivity growth. The quality of enterprises' innovation activity is rather more influenced by competition terms. Companies that are in the situation of moderate import competition achieve the overall increase in production output more often than others. At the same time, the increased competition with older Russian companies and, correspondingly, depleting potential of improving traditional products appear the significant factor of expanding new product output.

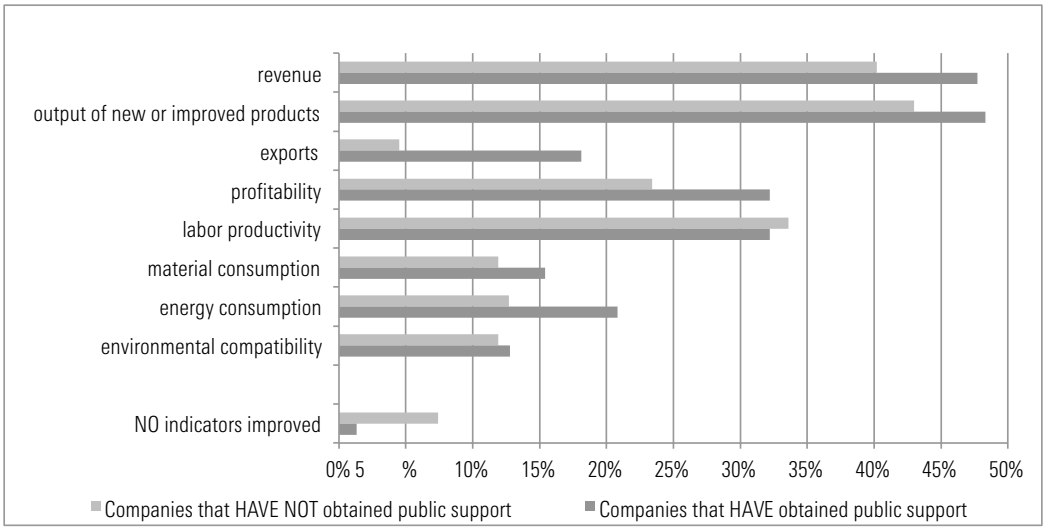


FIGURE 1 - Performance indicators increased due to innovations

3. WHAT FIRMS RECEIVE INNOVATION STATE SUPPORT MORE OFTEN: THE DISTRESSED OR THE DESERVING ONES?

Interpretation of the membership of state support benefaction does not in any case look like a trivial task. Notably, STI policy in Russia does not officially set any explicit goal in this field. So, what is better and more right in terms of community goals: to support a fairly effective firm by additional resources thereby ensuring the resources' efficient use, or allot funds to a firm that is just starting its innovation activity, switching it to the innovative mode, but with high risk of failure and loss of funds? Presumably, no unequivocal answer is possible in the system of rough evaluations available to us, but the question is still interesting: Is state support of innovation in Russia a bonus for leaders or a chance for stragglers?

In most cases the beneficiaries of the innovation state support policies are well-to-do companies, with up-to-date technology and exports. Therefore, state support for STI policy is oriented at the successful companies and not connected with the aid to outsider firms or those with partial state ownership as it was found by Garcia and Mohnen (2010), Lööf and Hesmati (2005), etc. Of course, this conclusion by no means excludes the possibility of distributing state support to ineffective businesses outside the innovation-fostering tools.

Generally, state support is allocated to relatively young companies, which may seem rather unexpected. Yet, this is a result of the existing structure of support mechanisms. The latter include the most massively applied depreciation bonus for investments in new equipment; notably, this tool is more important for startups intensively building their fixed assets.

It is interesting that the actively expanded use of tax incentives is largely due to the effect of competition with new Russian firms for the most favorable tax regime, while grants and substitutes are associated with increasing competitive pressure of imports. We assume that this reflects the government's attention to the subjects of preservation of jobs and protecting interests of national producers. These tasks are probably better solved by direct financial support from the state than anything else in the field of STI policy. Therefore, direct financial support for innovation is more connected with solving problems of business optimization for import substitution, whereas tax incentives are of greater assistance to business updating and development of new, globally-oriented companies.

4. MODELS OF BUILDING BUSINESS-STATE RELATIONS: SUPPORT IN EXCHANGE FOR THE CORRECT BEHAVIOR?

It was logical to suppose that the practice of extending state support to companies should have been related to the model of state-business relations. We also suggested that it is how friendly did the state treat state support recipients, how fair and how transparently did it select companies to be supported, and which were the important elements of creating motivation for conscientious companies to use state support mechanisms.

If the state does not attempt to manipulate business action, e. g. demand something from the business or constrain it, such business receives any state support for innovation significantly more rarely. Of course, one would think that the state limits public funding to the companies who fail to carry out some 'instructions', but it would be more correct to conclude by the set of attributes that companies outside the focus of the state try to 'stay unnoticed'.

At the same time, we failed to discover signs of exchange of businesses' social responsibility for the public support for innovation. Companies from whom the state expects social responsibility are less likely to become beneficiaries of state support for innovation. Moreover, initially we had proposed that businesses, which have a reputation of being socially responsible, would have more chances to receive subsidies due to the inevitable subjectivity of the decision-making; nevertheless, we received just the opposite result. This is possibly a manifestation of the desire to improve the uniformity and increase transparency in the state-business relations.

5. THE INFLUENCE OF STATE INNOVATION SUPPORT POLICIES ON COMPANIES' INNOVATIVE BEHAVIOR: IS IT POSSIBLE TO CHOOSE THE BETTER BETWEEN TAX INCENTIVES AND FINANCIAL SUPPORT?

Innovation state support mechanisms impact companies' innovative preferences and their choice of corporate priorities; these mechanisms also determine a set of specific behavioral changes in firms' innovative activity.

The analysis of the influence of support mechanisms on the companies' behavior has shown that the most common effect in Russia is crowding out of private funds by the public ones. This is due for both direct public financing and tax incentives. It seems reasonable to suggest that the demonstration effect of the state support for innovation has been rather weak, while for the beneficiaries of this support innovation is obviously not the only priority; sometimes it merely serves as a pretext for receiving additional resources.

The comparison of the influence of tax and financial mechanisms on companies' STI policy indicates that direct state financial support brings more fruit in the sense of launching new innovation projects, whereas application of tax instruments has a positive correlation with increasing the duration of projects beyond the three years or less project time stipulated by grant and subsidy schemes (Figure 2). This is consistent with findings of Guellec, Van Pottlesberghe (2003), Jaumotte, Pain (2005) and others. The observation that it is exactly the tax incentives and not the public subsidies that facilitate realization of long-term projects is likely a sequence of the flawed design of the standard Russian mechanism of public support, first of all, the short duration of project funding.

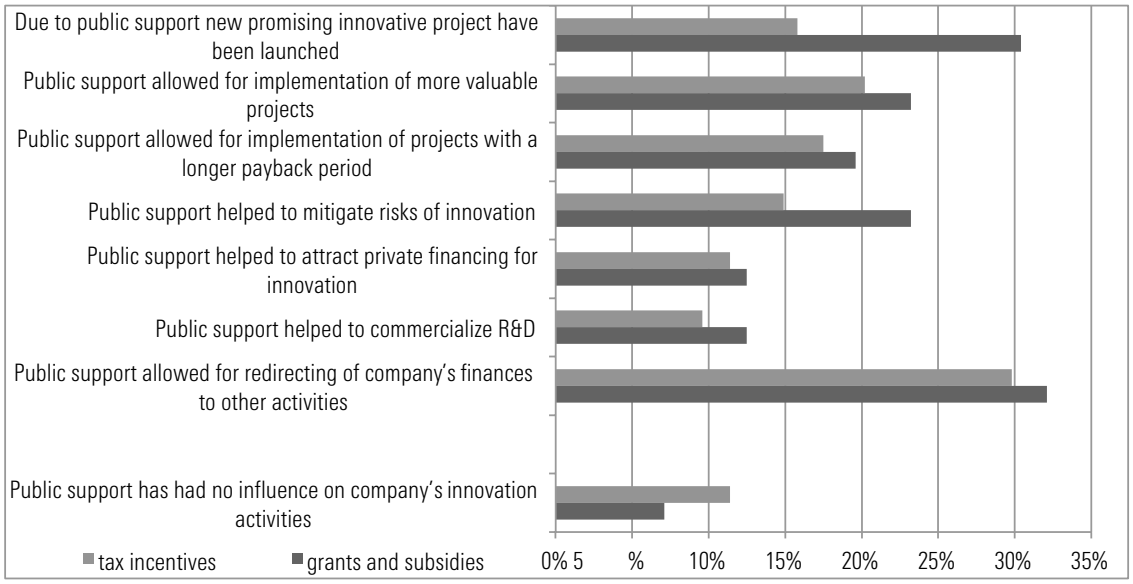


FIGURE 2 - Changes in companies' innovation activities due to public support

6. COSTS OF COMPANIES' ACCESS TO STATE SUPPORT OF INNOVATION AND RISKS OF ITS USE

Practicality and efficiency of state support policies largely depend on their main parameters (size and duration), terms of access, quality of administration, and ensuing risks of application for companies. Problems of access for companies and the risks of application seriously affect the effectiveness of the innovation policy (D'este et al. 2012). Figure 3 shows that the absence of sharing risks with the state has been most frequently cited as the most significant problem. In most cases this problem proved to scare away companies who had not used state support of innovation and, therefore, to hinder wider competition for obtaining state support (Garcia, Mohnen, 2012).

For companies who have received state support for their innovation activity, the most significant problem of tax incentives is the non-optimality of support parameters; public financing has suffered from red tape and complexity of the procedure of receiving support. Notably, both the tax and direct financial tools cause greater expenses for the young startups due to the requirements of additional paperwork and complicated accounting. That the design and administration of tax policies are more

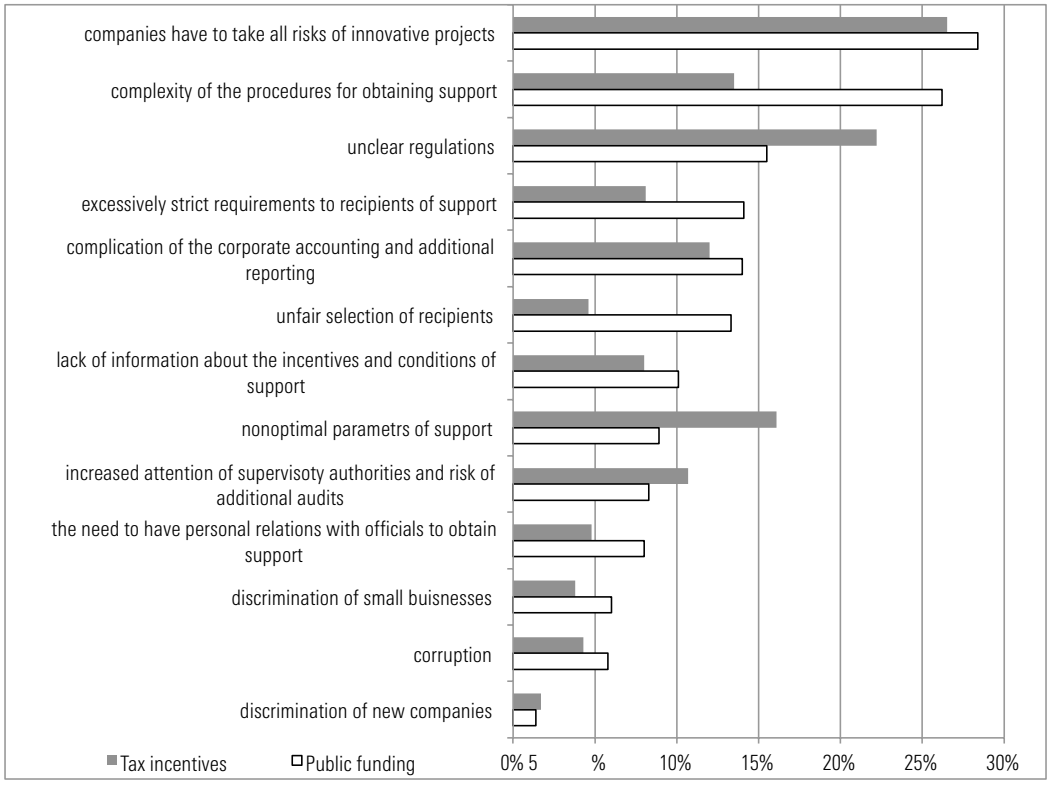


FIGURE 3 - Disadvantaged and problems of the STI policy instruments

friendly to older companies is a natural consequence of the fact that interests of already established business are easier to take into account; it looks more significant to authorities and its representatives have more developed connections with the government apparatus and have better opportunities to influence creation and adjustment of the support policies.

POLICY IMPLICATIONS

1. EVALUATING

The Russian practice of evaluating STI instruments is limited, not including the assessment of behavioral changes. As a result, even efficient incentive mechanisms sometimes look less advantageous because there is a lack of analysis of behavioral additionality.

The politicians are focusing on direct positive effects of the measures taken. The desire of rapid achievements leads to the risk of simulating positive results, when the officials rush to report about good results and consider the policy outcome as positive although it is not.

From various experiments in STI policy we need to move to comparative evaluation, portfolio evaluations and identification of good practices. The base for that is an independent assessment of effects. Moreover, the results of such assessments and the original data should be available to researchers affiliated with different interest groups.

2. STI INSTRUMENTS

No 'universally useful' innovation promotion mechanisms exist. When a new instrument is being introduced, it should first be applied neutrally and on a sufficiently broad scale; this will help to identify its real industry specificity and possible market failures, thus providing a basis for its specialization later on.

Counterbalancing the problems of the business environment directly by boosting stimuli for innovation is not a quite productive approach. On the contrary, such measures can further increase companies' motivation for rent-seeking behavior and imitating innovations.

It is necessary to develop an innovation-friendly regulation, and the government should truly share the innovation risks with businesses and has to be ready to lose some of its resources allocated for the support of innovations.

3. ACTIVISM AND STI POLICY IN THE WIDE SENSE

An excessively vigilant search for 'market failures' and the ways to compensate them may result in 'government failures' in this activity. This risk becomes even more significant if the government has limited ability to abolish unreasonable initiatives, especially in face of powerful lobbying by the traditional interest groups.

A considerable part of the barriers to innovation development of Russian economy is not directly linked to STI policy such as distortions in the competitive environment caused by different quotas and preferences; constraints on the growth of small and medium-sized companies; non-market benefits for some "socially significant" companies. These distortions reduce the demonstration effects of successful innovative companies, as well as the attractiveness of the relevant business behavior models.

REFERENCES

Bérubé C., Mohnen P. (2009). Are firms that receive R&D subsidies more innovative? CIRANO – Scientific Publication No. 2007s-13.

Bogetic Z., Olusi O. (2013). Drivers of Firm-Level Productivity in Russia's Manufacturing Sector. Policy Research Working Paper 6571, The World Bank.

D'Este P., Iammarino S., Savona M., von Tunzelmann N. (2012). What Hampers Innovation? Revealed Barriers Versus Deterring Barriers. Research Policy 41 (2): 482-488.

Dezhina I., Simachev Yu. (2013). Matching Grants for Stimulating Partnerships between Companies and Universities in Innovation Area: Initial Effects in Russia. Journal of New Economic Association 19 (3): 99-122.

Garcia A., Mohnen P. (2010). Impact of government support on R&D and innovation. UNU-MERIT working paper, Maastricht.

Goldberg I., Gobbard G., Racin J. (2011). Igniting innovation: rethinking the role of government in emerging Europe and Central Asia. World Bank, Washington DC.

Guellec D., Van Pottelsberghe B. (2000). The Impact of Public R&D Expenditure on Business R&D. STI Working Paper 2000/4, OECD, Paris.

Hanel P. (2003). Impact of Government Support Programs on Innovation by Canadian Manufacturing Firms. CIRST Working Paper.

Jaumotte F., Pain N. (2005). An Overview of Public Policies to Support Innovation. OECD Economic Department Working Paper No. 456.

Lööf H., Hesmati A. (2005). The Impact of Public Funding on Private R&D Investment. New Evidence from a Firm Level Innovation Study. MTT Discussion papers.

OECD. (2011). OECD reviews of Innovation Policy: Russian Federation 2011. OECD Publishing.

AUTHORS

YURI SIMACHEV
Interdepartmental analytical center
E: simachev@iacenter.ru

MIKHAIL KUZYK
Interdepartmental analytical center
E: kuzyk@iacenter.ru

VERA FEYGINA
Interdepartmental analytical center
E: feygina@iacenter.ru

EVALUATION THROUGH IMPACT: A DIFFERENT VIEWPOINT

EMANUELA REALE, MARIA NEDEVA, DUNCAN THOMAS and EMILIA PRIMERI

Paper presented at the Conference on "Evaluation of STI policies, instruments and organizations: New horizons and new challenges", Vienna, 14-15 November 2013

INTRODUCTION

Impact is an issue that has become really topical for policy makers in the last ten years. There are two main reasons to be interested in measuring and assessing the impact of policy and funding schemes: a) for the purposes of reflexivity and evaluation; and b) for the purposes of advocacy.

In terms of the former, the awareness of the impact of a particular policy or funding scheme, relates to what is considered to be the 'effectiveness' question. In the second case, the study of impact feeds into advocacy cases whereby some beneficial effects on reality (mainly in line with the state objectives) are claimed. In other words, the study of impact can be, and indeed is, used to evaluate and justify the continuation of particular policy and funding schemes in science and research.

The evaluation of impact traditionally faces the problem of attribution: the time-lag between the research result and the particular impact produced, as well as the fact that any impact has multiple causes, makes it difficult, or even impossible, to disentangle the causal relationship between the effect generated and the program under evaluation. Furthermore, assessing impact often lacks of adequate available information, data and indicators, hence the need of sophisticated models that merge qualitative and quantitative analyses (Mollas Gallart and Tang, 2011).

Other shortcomings related to the evaluation of impact are: a) the nature of impact differs according to the conceptions of the particular disciplines; b) the magnitude of impact can be different, and not all the impacts are neither desirable nor good; c) impact is often "indirect, partial, opaque and long-term" (Martin, 2011, 250).

In this paper, we argue that assessing the impact of research funding programs shall focus on unpacking the generative mechanisms that drive the programme design, the modes of application, on one side, and the representation that the mentioned design and practices produce on the beneficiaries' perceptions and research implementation, on the other side. Rather than looking at what concretely a research program has achieved as an impact, the proposed approach is interested in evaluating how far the program provided the beneficiaries with good opportunities for impact to occur. Thus, the ultimate aim of the evaluation is not to see whether a change happened because the programs produce the expected and desired effects, rather to improve the knowledge on whether the program created the right conditions for change, allowing the beneficiaries to do things better.

The proposed approach is discussed using the results from the impact evaluation of two research programs, one joint program and one open, both aimed at strengthening the internationalization of research, seeking for a more robust integration within the European Research Area (Nedeva, 2013).

BACKGROUND

THE NOTION OF IMPACT

There is a general agreement in the literature that the study of impact is gaining momentum and is becoming increasingly politically important (Kostoff, 1995; Donovan and Butler, 2007). It is probably fair to say that the importance of the study of 'impact' has been growing in parallel with the importance of policy. The study of 'impact' has been shaped by two different contexts. First is the increasing political demand for 'useful' research (Whitley 2010). This shift of public policies, which occurred after a period when research was funded for its own sake or on the basis of very diffuse promises of future uses, has also given rise to the search for assessment procedures that address the societal and/or economic impact of science (Kostoff 1995; Donovan and Butler 2007; Meagher et al. 2008).

A second context that promotes the study of impact is the imperative for evaluation of policy measures. The shift in public policies described above has led to the incorporation of public policy goals in many research policy measures, including those dedicated to the funding of research and innovations. The assessment or investigation of policy measures (or, more recently, of governance instruments) usually deals with these and other policies aimed at the promotion of contributions by science to societal welfare.

The underlying definition of impact as impact of science on other societal subsystems is in line with both general policy impact studies and the political interest in societal impact of research.[1] Hence, the notion of impact in this context is associated mainly with the economic and/or social effects that science may have, or it has indeed on society and economy.

1 '... A social impact is a significant improvement or deterioration in people's well-being or significant change in an aspect of community concern.' (Dietz, 1987, p. 56)

Another, more general notion of impact, incorporates the study of the effects that policies and funding schemes have on science itself (EU-RECIA Final Report). This generalised definition is in line with other generalised notions of impact, most notably the one by Becker (2001) who defines impact assessment as the process of identifying future consequences of current actions at individual, organisational or system level. According to this, impact could be defined as:

> *"...any difference and/or change of social actors or phenomena that can be partially or wholly attributed to other social actors or phenomena."*

The definition implies two key research tasks, namely identifying (measuring) change and attributing this to the funding or policy scheme that generates this change. Both tasks are non-trivial and entail a number of choices that frame and characterise the overall impact study approach as well as the choice of methodology(ies).

TYPES OF IMPACT

Generally speaking, evaluation is a method for understanding, on the base of perceptions, data, documentation, indicators, the results a specific activity achieved (program, treatment, drug, therapy, etc.), either they are desirable or not, temporary or permanent, immediate or long-term. When evaluation refers to a policy intervention, the objective is to formulate an appraisal that could be usable for governing the activities, driving the decision making by looking beyond the official goal.

Using motivations stated in the programmes' policy and funding objectives as points of reference for impact (also expected) to be generated by the programmes, four types of impact can be distinguished. These are illustrated in Tab. 1.

	INTENDED	UNINTENDED
EXPECTED	Straight runs	Collateral
UNEXPECTED	Long shots	Accidentals

TABLE 1 - Types of impact

Expectations regarding intended and expected impact ('straight runs') and intended and unexpected impact ('long shots') can be identified through the stated objectives of policy and research funding scheme. Whether or not these intentions are realised depends, on the one hand, on the support they receive by the core practices and on how clearly these are communicated and on the other hand, on how these are interpreted and used by the potential beneficiaries. *Whilst 'straight runs' are intended and anticipated, the 'long shots' are effects that are intended but cannot be expected to occur with any level of certainty within a set time frame.*

Unintended and expected impact ('collateral') is the 'collateral damage' that actors anticipate but cannot avoid because there are many social influences at play that the policy or funding scheme cannot control. Finally, unintended and unexpected impact ('accidentals') is very interesting as a possibility but difficult to measure. However, it can be captured if an empirical object is studied exhaustively. Most impact studies and assessments focus entirely on the 'straight runs' and 'long shots' types of impact.

TRADITIONAL APPROACHES TO THE STUDY OF IMPACT

The traditional approach to impact evaluation assumes one hypothesis on the effects that the policy (a funding program for instance) could have, and the changes it might generate. Three conditions must be satisfied: a causal event before the observed effect, a correlation between the two, and the fact that alternative hypotheses do not falsify the main one related to the impact. Experimental or quasi-experimental research designs can be envisaged (Herman, 1987) and different level of internal validity and statistical validity can be outlined, as well as different possibilities of generalizing the results. A further characteristic is that this type of analysis is not interested to understand how a specific impact occurred, rather to figure out robust analyses of linkages between the policy and the supposed impact.

We already mentioned that several problems rise when the impact assessment is concerned. They can be summarized in three broad types:

i. Methodological problems: measuring and attributing impact is problematic;

ii. Ontological problems: studies of impact assume that the relationship between the 'impactor' and 'impacted' is fairly direct and by necessity ignore the fact that in reality the social space is 'noisy' and there are many intervening factors/variables.

iii. Axiological problems: using result from the study of impact to evaluate the performance of, and decide the destiny of, policy and funding schemes is problematic since the actual outcome has only very distant relationship to signals and actions from the scheme.

CONCEPTUAL FRAMEWORK

The approach prosed in this paper is based on the two major rationales that, in our view, can explain the specific aim of research funding, e.g. internationalisation. One is 'normative' (achieving a political aim), the other is 'problem solving' (i.e. addressing issues faced by knowledge dynamics). There are at least two consequences for the overall analysis deriving from these remarks, which also let emerge several questions:

a. In the policy-oriented framework, what is at stake is a measure of dynamics at work. Is internationalisation growing? At which speed? Does it extend research fields? In which fields is this more intense? Which are the most productive instruments? What explains their uptake?

b. In the knowledge dynamics framework, the focus is on: What triggers the need to move to the EU level (science dynamics per se, technology and industry arrangements, societal issues)? Is this a shared feature, or does it relate to different national situations (mostly in terms of critical size, but not only)? What are the processes through which these objectives are framed and implementation structures created (in particular what explains the selection of particular instruments)? What are the institutional constraints /

levers for such evolution? (Is it easier in "euro-compatible" institutional arrangements, in which the intermediate layer is made of strong and independent (quasi) funding agencies?)

Each research funding programme can be seen as a mechanism embodying four kinds of opportunities, namely 'intended opportunities', 'provided opportunities', 'perceived opportunities' and 'mobilised opportunities'.

The 'intended opportunities' are linked to policy rationales, which can also be re-constructed either by reading the missions and objectives of funding mechanisms or through a historical reconstruction of the 'conditions of origin' of the mechanisms. This is far beyond to the evaluation of impact as related to the official goal of the program; rather it is interested to look to the consequences the programs is likely to produce (Weiss, 1993). Intended opportunities impact the capability to set out the right goals and objectives improving the reflexivity of the policy action.

Whether or not the intended opportunities have been/are being provided by a programme depends on the way in which they are implemented through three different practices: selection practices, funding practices, and accountability practices. It is important to mention that the funding mechanism (programme) can be expected to achieve its 'intended' impact (effects) if there is a high level of congruence between 'intended' and 'provided' opportunities. For instance, a programme aiming to offer opportunities for technology transfer and using selection, funding, and accountability practices that do not match this intention is likely to have effects that are different from the intended ones. A particularly tricky point is accountability since 'what is counted is usually delivered'.

The beneficiary groups perceive each research programme (or funding opportunity) as a particular kind of opportunity. The kind of opportunity it may be perceived mainly depends on two things: a) the way in which the programme is presented; and b) the opportunities that potential beneficiaries consider important for them since the possibility for a program to produce an impact has to take into account the values of the epistemic communities. Thus, provided opportunities impact the decision of the potential beneficiaries to apply or not.

Potentially, the biggest mismatch can be expected between the opportunities that are intended and provided, on the one hand, and the opportunities as they are perceived, on the other hand. The bigger this mismatch, the higher the probability that the programme is not going to achieve its 'intended' effects. Detecting and explaining said mismatch allows researchers to explore and map different 'unintended' and 'unexpected' effects that a funding mechanism may have.

Opportunities can be analysed not through the intentions and interpretations of the beneficiaries but through their actions. In other words, this approach looks at what has been/is being achieved for what concerns participation in the programme and what the beneficiaries see as their main achievements.

Mismatch might occur in this case too –beneficiaries may perceive the opportunities that a funding mechanism provides but decide that these are not the opportunities they need, so they 'bend' them to suit their needs. An example regards some of the funding from the ERC (EURECIA, 2012, Nedeva, 2013): a grantee said that he recognised that the funding was important for basic science and path-breaking research, but for him the way to mobilise this opportunity was to use it to buy equipment, because it was what he needed and other sources were much more restrictive. Mobilised opportunities refer first and foremost to the expected benefits that programmes are supposed to produce; the impact occur on further research activities developed by the epistemic communities.

Links with intended and provided efforts can be detected in the motivations for further action, which might produce a feedback effect (Fig. 1).

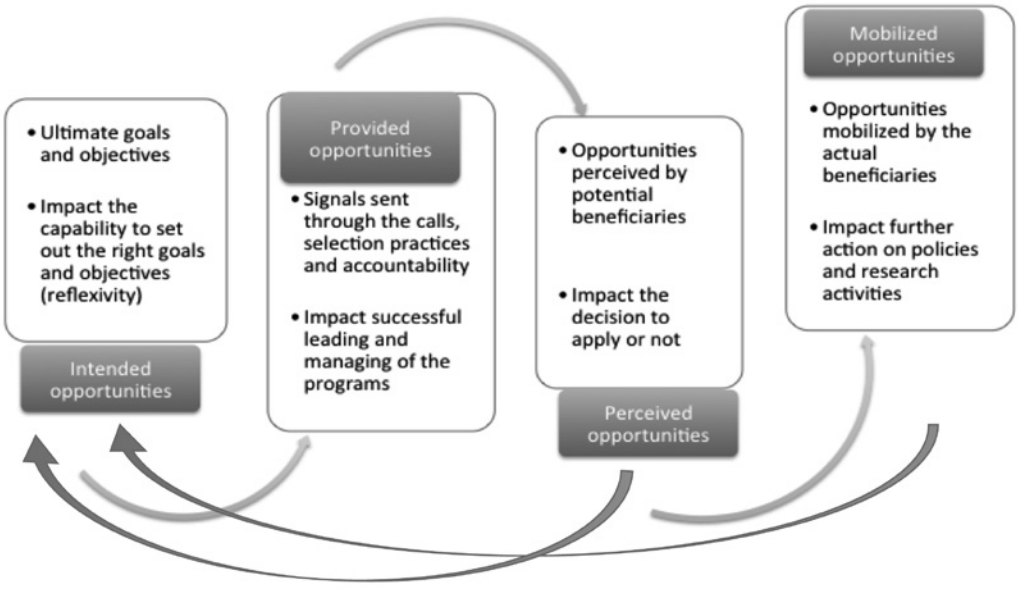


FIGURE 1 - Programs as set of opportunities

2 JOREP Project. The countries involved are: Czech Republic, Denmark, France, Germany, Italy, Norway, Poland, Spain, Switzerland, The Netherlands, UK.
3 ORA at the moment includes DE, FR, NL, UK, US

The paper uses two out of ten case studies from a project developed under the auspices of the European Commission on joint and open research programmes developed in eleven European countries.[2] We refer to SINERGIA in Switzerland and Open Research Area (ORA) -which at the time of the investigation joined France, UK, NL and Germany[3], for testing the proposed approach of impact assessment. They are both national–based funding schemes aimed at strengthening the internationalization of the national research system either by allowing the possibility to include one non-national team among the funded participants or by creating a specific joint mechanism between Funding Agencies of different countries (Reale et al, 2013, Lepori and Reale, 2012).

Based on the conceptual framework, the methodological choice of the paper for impact assessment is presented in Fig.2.

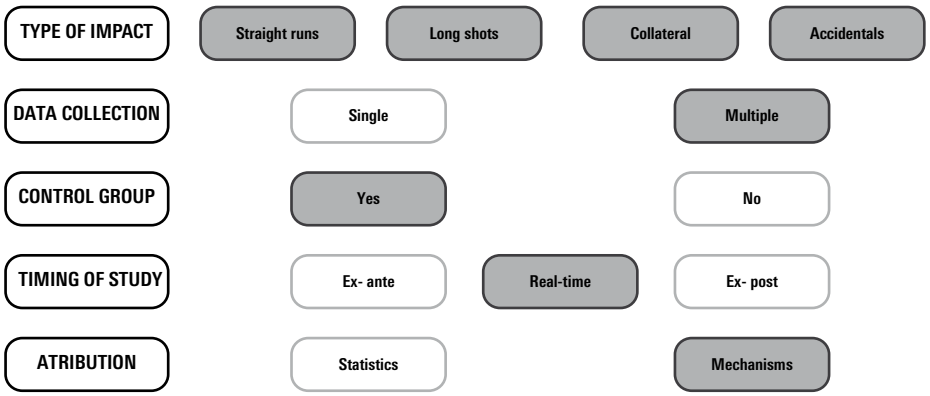


FIGURE 2 - Methodological choice (Choices grey shadowed)

The case studies include a documentary analysis, an empirical analysis supported by semi-structured interviews (5) to the programmes officers, and a survey. The overall response rate to the survey was 74 individual beneficiary responses out of an original sample of 127 people. The population only includes those that have been funded by the programmes; no information was collected on applicants not selected for funding.

The analysis is aimed at unpacking the unique opportunities perceived by the epistemic communities as to: a) *the rules and management of the programmes* (more money for research, a longer funding duration, less bureaucratic fulfilments, a greater fairness and transparency of the selection and evaluation processes), b) *the advantages for the research activity with respect to other existing funding schemes* (a bigger partner networking, geographic and/or intellectual cross-boundaries, more robust industry-academic collaboration, risk-taking activities, wider internationality of the partners), and c) *the exploitations of the results* (a better dissemination into a wide audience and a more efficient intellectual property rights management). The questionnaire asked for each item whether the opportunity has been perceived before the starting of the application.

The results (Tab. 2 and 3) confirm that the programs have generally met the participants' expectations as to perceived and mobilised benefits (YY-[4] answers); a rather large share of respondents realised unexpected opportunities (NY), while the number of beneficiaries whose expectations were not satisfied is relatively small (YN). Our exploration shows that SINERGIA has been strong enough and the signals were unique enough to be able to bring about changes in the nature of knowledge and productivity through its grants. The analysis of ORA does not allow for the same conclusions. The program is very useful as part of a set of policies, but has have little influence individually.

	RISK-TAKING	BUREAUCRATIC ASPECTS (LESS BUREAUCRACY)	INDUSTRY- ACADEMIC COLLABORATIOR	TRANS-NATIONAL/CROSS DISCIPLINARY RESEARCH
YY	29,5%	55,3%	0%	74,5%
NY	40,9%	44,7%	4,4%	14,9%
YN	0%	0%	2,2%	8,5%
NN	29,5%	0%	93,3%	2,1%

(% of respondents = 47 out of 85) 63,3% cross-disciplinary, 30% trans-national

TABLE 2 - Perceived and mobilized opportunities – SINERGIA

4 Four groups of answers according to opportunities perceived/mobilized: YY = yes-yes ; NY= no-yes; YN=yes-no; NN=no-no

	RISK-TAKING	BUREAUCRATIC ASPECTS	INDUSTRY- ACADEMIC COLLABORATIOR	TRANS-NATIONAL/CROSS DISCIPLINARY RESEARCH
YY	23,8%	37,5%	0%	88,5%
NY	9,5%	58,3%	4,3%	7,7%
YN	9,5%	0%	0%	0%
NN	57,1%	4,2%	95,7%	3,8%

% of respondents = 27 out of 42; trans-national research 91%
TABLE 3 - Perceived and mobilized opportunities – ORA

CONCLUDING REMARKS

The paper presents a point of view for dealing with the evaluation of impact, which assumes that the changes a research-funding scheme is suitable to produce in science, economy and society are out of the control of decision makers. Beside attribution (either 'sole' attribution or contribution, White, 2010), ontological and axiological problems constrain the possibility to capture the impact.

The proposed approach shifts then the emphasis from impact evaluation to whether the program is able to create the conditions for change; the focus is to assess the possibilities for an impact to enter in action, thus what makes the funding scheme unique or substantially different from the others. According to this perspective, assessing the success and effectiveness of a program is not a matter of measuring its impact; rather it relies on the evaluation of the impact that the program is potentially suitable to produce (that is measuring its effectiveness). Intended opportunities impact on the capability of policy makers to set out the right goal and objectives; the provided opportunities impact on the leading and managing of the programs (successful engaging in decision-making activities); how opportunities are perceived impact on the decision of the scholar communities to apply or not, together with the internal appreciation of the scholars (a fact that must be taken into account for reflexivity on the policy action), and the mobilized opportunities impact on further action on policies and research activities, with a clear possibility to produce a feedback on intended opportunities. Thus, the content of evaluation is not how far the programs 'do the better things', rather how far the programs allow 'to do things better', addressing questions and challenges that could not be addressed otherwise.

The limitations of the explorative test proposed depend, on the one hand, on the number of cases analysed, and, on the other hand, on the fact that impact evaluation needs to go more inside the functioning of the mechanisms, in order to better understand how the signals they provide are generated, and under what conditions they are likely to produce unintended events.

Despite the mentioned limitations, the approach looks promising; both cases show the concrete ways in which the programs have been implemented, and the good practices (such as transparency, flexibility, and reliability of the selection process) improving the possibility of a program to be successful. More research on the generative mechanisms has to be carried out in order to figure out impacts on the content of research (new lines under risk-taking), the impact on funding agencies (flexibility, bureaucracy and funding portfolio), on institutions and individuals (as to the issues of internationalization and cross-disciplinary research). Investigating unintended –either collaterals or accidentals, is the most challenging possibility to be tested against the different point of view the paper proposed.

REFERENCES

Becker H. A. (2001). Social impact assessment, European Journal of Operational Research. 128, 311-321

Donovan, C. and Butler, L. (2007). Testing Novel Quantitative Indicators of Research "Quality", Esteem and User Engagement: An Economics Pilot Study. Research Evaluation, 16(4): 231-242

EURECIA (2012). Understanding and Assessing the Impact and Outcomes of the ERC and its Funding Schemes. Final Synthesis Report, European Commission, May, available at: http://erc.europa.eu/sites/default/files/document/file/eurecia_final_synthesis_report.pdf

Lepori B, Reale E. (2012). S&T indicators as a tool for formative evaluation of research programs. Evaluation, 18(4), 451-465
Kostoff R. N. (1995). Federal research impact assessment: axioms, approaches, applications. Scientometrics, Vol. 34, No. 2 (1995) 163-206

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

Meagher L. et al. (2008). Flows of knowledge, expertise and influence: a method for assessing policy and practice impacts from social science research. Research Evaluation, 17(3), 163-173

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

Nedeva, M. (2013). Between the global and the national: Organising European science. Research
Policy, 42, 220-230

Reale, E., Lepori, B., Nedeva, M. & Thomas, D. Primeri, E., Chassagneux, E., Laredo, P., (2013). Understanding the Dynamics of Research Funding Collaboration in the European Research Area. JOREP Final Report. European Commission, Luxembourg

White H. (2010). A contribution to current debates in impact evaluation. Evaluation, 16(2), 153-164

AUTHORS

EMANUELA REALE
CERIS CNR - Institute for research on firm and growth, National Research Council
E: e.reale@ceris.cnr.it

MARIA NEDEVA
MiOIR-Manchester Business School, University of Manchester
E: maria.nedeva@mbs.ac.uk

DUNCAN THOMAS
MiOIR-Manchester Business School, University of Manchester
E: duncan.thomas@mbs.ac.uk

EMILIA PRIMERI
CERIS CNR - Institute for research on firm and growth, National Research Council
E: e.primeri@ceris.cnr.it

MONITORING AND EVALUATION IN JOINT CALLS OF "HORIZONTAL – INCO" ERA-NET AND ERA-NET PLUS ACTIONS

MARTIN FELIX GAJDUSEK, NIKOS SIDIROPOULOS

INTRODUCTION

With the emergence of ERA-Nets in FP6 and its continuation and extension to ERA-Net PLUS in FP7, the EC framework programmes stimulate pooling of national research funding in thematic fields (thematic ERA-Nets) or support joint calls with third countries (INCO ERA-Nets). While in FP6 the dominant mode of establishing an ERA-NET was a bottom up approach for the specific theme covered by an action, the thematic topics in FP7 were almost exclusively nominated top-down by the programme committees of the thematic directorates of the EC. The most established ERA-NETs in FP6 focused on analysis and consultation in order to prepare joint calls. The joint calls in all ERA NET projects were supported by the project partners and other financing partners beyond the partnership with the aim to pool funding. In a bottom up process the networks searched for commitment from funders, different funding modes and rules were established to operate these funds. Beside the financial commitment of the funders, the non-negotiable framework conditions predominantly stemming from the provenience of funding had to be taken into consideration. Other substantial factors to be considered were the scheduling of calls, the just-retour principle, the acceptable final beneficiaries of funding summarized in national eligibility rules, specific spending rules and a number of other framework conditions for setting up a joint call among others.

Some of the established ERA-Nets succeeded to set up pilot joint calls or regular calls, in the best case following up their activities from FP6 as well in FP7. Others established successful Art.185 initiatives (like e.g. BONUS), and some used ERA Net PLUS to continue joint funding with stronger EC funding involvement.

The main difference in FP7 was the general thematic top down approach and the introduction of the ERA NET PLUS instrument that involves directly the EC in funding activities to third (final) beneficiaries. A top up for a single joint call was provided in this case. The key difference of horizontal "INCO" ERA NETs addressing particular regions or single countries is the unchanged bottom-up approach as the joint calls are obviously still considered to be pilot activities although some of the pilot calls have established rather mature procedures.

ASSESSMENT AND MONITORING OF THE INSTRUMENT

The implementation of the ERA-Nets and their joint calls, their volume and frequency is monitored by the JRC while the scheme is also monitored by DG Research and Innovation. Annual joint programming events inform about ongoing developments and support information and exchange with the main partners of ERA-Nets, namely programme owners as well as agencies and ministries.

The monitoring of the instrument was planned systematically by the JRC/IPTS and the analytical framework is publicly available[1]. The ERA-NET coordinators were approached by a number of surveys collecting systematically data about the progress of the instrument. Base data for each established ERA NET are available on the Netwatch[2] portal. Also regular update requests are addressed to the coordinators of ERA NETs. In addition to the ERA NET the ERALEARN[3] platform supports learning by providing practical information for the funders or prospective funders to support transnational calls in a number of aspects e.g. through newsletters, relevant EC communications, workshops etc. Directorate B monitors the progress of the instrument as well and provides key findings in reports[4].

On national level research councils, national ministries and agencies are highly interested in the performance and efficiency of the instrument. Exemplary, the Austrian PROVISO monitors since many years the EU-FP´s and produces a national monitoring of ERA NET involvement[5].

As a matter of fact the core activity of ERA-NETs, the pilot calls, are rather a complex joint effort and it is difficult to understand the effects of funding on the level of third beneficiaries, the final recipients of funding distributed through the joint calls. The programme owners and agencies

1 Susana Elena Pérez , Hans-Günther Schwarz: Developing an analytical framework for mapping, monitoring and assessing transnational R&D collaboration in Europe. The case of the ERA-NETs. (2009) JRC, IPTS Scientific and Technical Research series – ISSN 1018-5593; DOI 10.2791/11940; ISBN 978-92-79-12562-1

2 Several categories of data can be accessed on: http://netwatch.jrc.ec.europa.eu/web/ni/network-information

3 ERA LEARN (now ERALEARN2) is now closer linked to the NETWATCH of IPTS/JRS; http://www.era-learn.eu

4 As an example the document http://ec.europa.eu/research/era/pdf/era_net_report_statistics_2013.pdf provides important information e.g. on progress of the instrument and volume of funding, frequency of calls etc.; information is widely based on results produced by the IPTS/JRS Netwatch.

5 http://www.bmwf.gv.at/home/bmwf/research/the-proviso-project

involved are acquainted with the process of joint funding on national level, the procedure established through joint call secretariats and the distribution and use of committed funding to joint calls.

The opposite is the case with the knowledge about the effects on the final beneficiaries, the effects of international cooperation and the added value. The knowledge about these effects is rather limited and seems to be not systematically evidenced. A better understanding of the practical dimensions of transnational funding could help to establish potentially more efficient or more relevant and far reaching joint funding activities.

thematic ERA-NETs are used as a comparison group. The methodological challenge to establish direct matching pairs would be size of funding, maturity of network, similar country coverage, thematic coverage and other factors that make an adequate comparison almost impossible. Still, in order to understand whether monitoring processes in thematic ERA NETS are mature in comparison to the INCO ERA NETs, a few thematic ERA NETs were included in the analysis[8].

According to our understanding the main differences between INCO ERA NETs and thematic ERA NETs would be as follows (see Tab. 1):

Horizontal INCO ERA NET calls	Thematic ERA NET calls
• INCO Policy driven • Joint identification process of topic(s) • Regional/national priorities • Pilot character • Smaller budgets • Funders' activities' portfolio is broad • International cooperation departments of ministries, rarely agencies • Bundling bilateral funding • Excellence criteria + Policy driven action • EU MS effort + country/region effort • Variety of funding rules beyond EU MS • Importance of EC labeling	• Indirectly policy driven • Given thematic focus + fine tuning • Limited regional scope • Maturity: Repeated, multiple calls • Substantial budgets • Funders rather focused on funding • Agencies of thematic programmes, partly owners • Bundling funding to reach critical mass • Excellence, • EU MS funding is driving force • EU MS funding rules (exceptions)

TABLE 1 - Differences between INCO ERA NET calls and Thematic ERA NET calls

The INCO ERA NETs as the key object of investigation of this paper address a particular region or a third country to the EU FP. They are generally policy driven[6] and define together with the countries involved the joint call priorities. Moreover, they involve programme owners in charge of international R&D cooperation rather than agencies, operate smaller budgets and possess experience from bilateral co-operation. Furthermore the expectations are not mainly directed to excellence, but rather focus on inclusion and linking-up of research communities. Obviously, the expectations of "target countries/regions" (from EC perspective) and the EU MS involved can differ widely.

The authors have been involved in horizontal "INCO" ERA Nets and have set up or conducted monitoring tasks in order to facilitate understanding of impact-dimensions of joint funding, to support future streamlining of funding. Practical experience with the instrument allows an insight in the potentially triggered effects one can expect. Exploring the potential and effect of INCO ERA NET joint funding needs more profound involvement comparing to thematic ERA-NETs[7]. In the current paper the authors discuss the INCO ERA NETs and aim at exploring the current monitoring processes established, the potentially planned impact evaluations and their feasibility. In order to maintain a broader perspective,

On one hand, impact dimensions concerning the funding parties and the EC itself, this would be e.g. institutional learning, enhanced visibility, justification of public spending, interaction with other funders and a number of other dimensions, are all subject of intensive research work and monitoring – predominantly but not exclusively - by the JRC.

On the other hand, when looking on the impact dimensions on final beneficiaries of a call e.g. the interaction of individuals and host organisations, their outputs and learning, additionally invested resources, additional outputs, and networking effects could be listed as most prominent[9]. These effects are widely unexplored in the context of horizontal INCO ERA-NETs.

The authors of this paper address the following research questions:

- Do systematic assessments of call outputs and results exist? How is the evaluation culture of call outputs?
- Are there clear "programme" objectives reproduced in the call?
- Would clear objectives allow (better) monitoring and evaluation processes?
- Do any indicators exist to measure success or impact?
- Is there any prevision of impact evaluation in future?

6 The policy dimension addressed here could be defined as "international S&T cooperation policy", that differs from the overall thematic ERA-NET policy goal of pooling funding and coordination of funds of EU MS.

7 The main difference can be found along the policy dimensions for the EU, the declared interest to work with a specific region.

8 For a full coverage study on all ERA NETs matching criteria for a comparison must be defined more carefully e.g for thematic similarity, involvement of 3rd countries in thematic ERA Nets, the eligibility criteria for the type of beneficiaries, the funding mode and role of call secretariats etc.

9 Additionality would be one concept that could help to understand the processes at final beneficiaries.

In addition we emphasize on the feasibility and preconditions for advanced monitoring and evaluation tasks in future.

The first question addresses the distribution of work and the information needs, e.g. if the monitoring of funded activities – if at all - is done by a call secretariat, by the funders themselves or externals employed. The second question concerns the existence of clear objectives of a call and the existence of success indicators. The ultimate question is, how monitoring has been structured and whether (impact) evaluation would be feasible in future. In addition, the question of the usefulness of evaluation in its learning function should be addressed as some of the calls are one-off activities. Here the question is posed whether documentation on final beneficiary level is perceived as important and as an essential element of the transnational funding activity.

The IPTS/JRS with its analysis provided through NETWATCH, indirectly ERALEARN, emphasizes on the highly relevant policy and strategic dimensions. Our analysis though looks in more in detail on the monitoring efforts directed to the final beneficiaries of calls.

The authors addressed an online survey to eight INCO-ERA-NETs and four thematic ERA NETs to explore the current and potential monitoring processes, the feasibility of summative evaluations, the overview on existing evaluation knowledge (see Tab. 2)[10].

Target country/region	Acronym	Calls: Funding completed	Calls: Open call or Ongoing Funding	Calls: Planned	M€ Committed [1]
India	NEWINDIGO	1	1		6.82
Japan	CONCERT JAPAN		1		5.29[2]
Korea	KORANET	1	1		2.93
Russia	ERA NET RUS	1	1	1	24.84
Africa	ERA AFRICA		1		1.07
Black Sea Countries	BS ERA NET		1		3.50
Latin America (FP6)	EULANEST	1			1.57
Western Balkan Countries	SEE-ERA.NET PLUS	1			3.41
Target country/thematic ERA-NET	**Acronym**	**Calls: Funding completed**	**Calls: Open call or Ongoing Funding**	**Calls: Planned**	**M€ Committed [1]**
ICT+Agriculture (Israel)	ICT AGR(incl. -2)	1	1	4	9.33[3]
Forest research (Algeria, Morocco, Tunesia)	FORESTERRA		1		1.50[4]
Biotechnology (Russia in 8th call)	ERA NET EURO TRANSBIO	4	3	5	176.19[5]
Israel, Canada, US (parallel US-NSF call)	ERACAPS		1	1	20.00

[1] See Annex I (Sept 2013) in: Bertrand, E. and Niehoff, J.; Report on ERA-NET, ERA-NET Plus and JPIs and their joint calls; European Commission (2013)
[2] http://www.concertjapan.eu/system/files/CJ%20JC%20Call%20Text%20and%20Regulations_Final.pdf last accessed on 1/11/2013
[3] Excluding upcoming funding in follow up project ICT-AGR-2
[4] Recent launch of call
[5] Excluding the 3 calls planned in a follow up project to the ERA-NET

TABLE 2 - ERA NETs under scrutiny

10 No existing project covered the costs of this paper. The online survey was the only available method beside online resources like the call documents, ERA NET websites etc. We are aware that it would have been an advantage to employ also interviews as the topic is complex.

THE CLEARNESS OF OBJECTIVES OF ERA NET CALLS

One idea of ERA-NET calls is the pooling of existing research funds. Netwatch provides information about the provenience of funds. Not surprisingly **new** funding was directed in the INCO ERA Nets, funding that is regularly administered by the ministries, funds are not compulsory part of existing programmes and have not been handed over to agencies. Given that situation, own (national) goals and objectives, compulsory monitoring and evaluation procedures that maybe exist in national programmes do not exist for funding dedicated to international cooperation.

Additional policy driven objectives along the "internationalisation" dimensions are rather in the centre of setting up joint calls. These are based on analytical work conducted during the ERA-NET. These observations lead to the assumption that the established calls can be treated in the analysis as a new "programme" set up by the funders[11].

One would expect that clear programme documents defining goals of (a) call(s) would exist when joint funding is set up, containing the call objectives, quantifiable targets etc. . According to our understanding the following processes replace such "programme document" in ERA NETs:

- Announcement of ERA NET by the EC according to work programmes (indicated funding volume, call volume, duration of EC support etc.);
- Application to the EC including the commitment demonstrated[12], procedures described, indicating call topic(s) and demonstrating the overall impact of the activity;
- Call documents describing the expected cooperation goals, including annexes on funding limits and national funding rules.

One could add to the list the strategic research agendas established in Art.185 initiatives, as they are also employing ERA-NETs for funding with a dedicated regional coverage (e.g. BONUS).

An interesting topic for further investigation would be the question whether the number of documents can replace a single "programme" document.

The lack of description of objectives in a clear programme document significantly affects the definition of selection criteria for calls, and the complex linkages of selection criteria to specific call objectives challenge potentially evaluators looking on the impact and relevance of the funding activity.

In the case of ERA NET PLUS the selection criteria are defined with Annex III of the contract with the EC, so that a group of funders does not have any choice to establish better fitting criteria that one would expect in rather policy driven actions like in an INCO ERA NET PLUS.

THEMATIC PRIORITY SETTING AND THE OBJECTIVES OF THE CALL

One of the specificities of pooling funding is the work-share employed by the funding partners. Joint call secretariats are established that also set up a draft call document further negotiated with the funding partners. Main points of negotiation are the schedule of calls, the national funding rules and processes to set up clear implementation guidelines according to the funding model employed[13]. A number of negotiables and non-negotiables influence the final call document released.

Based on the general typology of horizontal and thematic ERA NETs one can derive main differences for the two types. Table 3 compares the core objectives and differences as far as they are relevant for monitoring and evaluation of calls.

For the policy driven objectives the INCO ERA Nets have some particularities; they aim at facilitating access, networking and intraregional cooperation. The motivation of funders as summarized in the second column shows also the link with policy actions, the emphasis on synergies of cooperation and the thematic needs of countries and the target regions. The definition of call topics "against background of priorities on policy level" is also a particularity of the INCO ERA NETs[14].

Provided with the analytical evidence gathered in horizontal ERA NETs one would assume that it would be feasible to define clear call objectives in a call document. Evidently the policy driven objectives are presented in call documents, but quantified indicators are not part of call documents[15].

Table 4 compares the set ex-ante selection criteria that take up specific objectives as published in the call documents.

The main categories in the first column show the different way of expressing excellence, economic and social impact, management and planning, and the way of implementation proposed. One must acknowledge that some of the selection criteria in call documents are standar-

11 It would need more definition of "programmes" and the difference to transnational joint calls, even if repeatedly implemented or functioning without EC involvement that could be objected. Some substantial funding can be observed in thematic ERA NETs that continued without EC funding, here the term "programme" would be more applicable.

12 The Netwatch information provides evidence on funding committed, also the assessment of the Directorate B on http://ec.europa.eu/research/era/pdf/era_net_report_statistics_2013.pdf provides evidence on the number of calls planned in ERA NETs. Committed funding is demonstrated with "Commitment Letters" that do not necessarily mean that distributed funding must reach the indicated amount.

13 This article does not emphasize on the funding mode employed, but assumes that in INCO ERA NETs the juste retour principle is the key concept employed by all funding activities.

14 Policy driven objectives are not prominent in thematic ERA NETs that are considered to be slightly faster than INCO ERA NETS with launch of calls, but no general statement can be presented given the small sample of thematic ERA NETs approached.

15 This would be "number of projects supported in call" or more specific output definitions like peer reviewed publications cited in WoS or SCOPUS, networks established, etc. - any other quantifiable output expected for the whole call.

	Policy driven objectives	Motivation of Funding partners	Set up of call objectives Timing, involved groups	Definition of Call topic
INCO ERA NETs	Access Networking Intraregional cooperation Pooling bilateral funding Critical mass funding Piloting/Feasibility Sustained cooperation Thematic cooperation	Linkage with policy actions Sustainability of previous ERA NET EC top-up in ERA NET Plus Synergies of cooperation Openness of instrument Thematic needs	1 year in advance to calls FP7 application (ERA NET Plus) Task leaders, (funding) partners Thematic consultation	Analysis Thematic workshops Against background of priorities on policy level Current event (i.e. the Tsunami event in Japan) EU MS Funding partners All Funding partners
Thematic ERA NETs	Excellence Outreach to SMEs/ collaborative research Transnational co-operation of SMEs	Visibility of thematic cooperation across EU Coordination to avoid duplication Input from other sources of complementary funding Learning among researchers	3-6 months in advance to calls Application to ERA NET Consultation with stakeholders Internal meetings, Management Board External High Level Group	Open call in thematic ERA NETs Community asked Workshop of funders and potential beneficiaries Scientific Advisory board + High Level Group

TABLE 3 - Priority setting and objectives of call (analysis from survey answers)

	MAIN CATEGORIES	ADDITIONAL CATEGORIES
INCO ERA NETs	Scientific Excellence Expected impact Quality of consortium Quality of project management Quality of the demonstration of proposal	Regional focus Transnational added value Sustainability Exploitation of results Originality, novelty Multi/interdisciplinarity Involvement (Gender, young scientists)
Thematic ERA NETs	Technological and scientific excellence Social and economic perspective Consortium and Project Management Resources	European added value Transnational added value Exploitation and IPR Networking

TABLE 4 - Ex-ante selection criteria (analysis from survey answers and call documents)

dised copy-paste solutions[16]. When looking on the horizontal ERA NETs the clear notion of the regional focus, sustainability, interdisciplinarity and the involvement e.g. of young scientists or gender equality are additional categories for the assessment of proposals. Interestingly the European added value was not pointed out in the INCO ERA NETs, while they are considered as "efforts of the Member States" by the EC[17].

A set of questions was addressed to the ERA NETs that concern the monitoring processes of funded projects. We distinct content related monitoring and financial reporting monitoring as implemented through audit processes. The scheduling of monitoring or evaluation, interim or at the point of termination of funding, was explored with the survey. As shown in Table 5, financial processes are clearly defined and procedures seem to be adequate, while content related documentation is employed mainly in the course of interim reporting. A single ERA NET[18] has employed all feasible monitoring and audit steps according to information received from the survey.

Some employ also interim monitoring meetings to assess progress of work. It would be interesting to look deeper in the typology of interim meetings to fully understand their function.

As shown in Table 6, a number of currently running ERA NETs plan impact evaluations. The figure shows together with Table 5 the preparedness for impact evaluations[19].

Considering the above figures one could ask how the *planned impact evaluations* can be conducted sufficiently at all if e.g. there are no sufficient procedures to collect outputs and results of joint work, or if no access to the current beneficiaries of funding will be guaranteed[20]. As a whole,

16 One must acknowledge that the German DLR is a main provider for the administration of horizontal ERA NET calls and has set up a number of call documents. E.g. the "Quality of the Demonstration of Proposal" is a specificity of the INCO ERA-NET call documents and was repeated several times but not any similar category can be observed in thematic ERA NETs.

17 Maybe it is uncomfortable to show that position to a third country addressed.

18 ERA NET RUS focusing on Russia

19 However, one ERA NET declares according to Table 6 to collect systematically all outputs and results (column 2) while not having established reporting procedure on content matters as shown in Table 5.

20 One of the analysed projects can be quite confident about the future access to beneficiaries but has not planned an impact assessment (or evaluation).

Content			**Financial**			
Interim	*Interim*	*Final*	*Interim*	*Interim*	*Final*	*Final*
Content reporting	Monitoring meeting	Content reporting (targets)	Financial reporting to CS	Financial reporting to Funding Partners	Financial reporting to CS	Financial reporting to Funding Partners
X			X		X	
X	X		X		X	
X				X		X
X		X		X		X
X	X	X		X		X
				X		
X	X	X	X	X	X	X

Note: thematic ERA NETs not displayed
Note: "X" indicates the existence of procedure. CS: Call Secretariat
Source: Own data from survey to ERA NETs

TABLE 5 - Monitoring procedures in INCO ERA-NETs

Impact assessment planned	**Effort to collect all projects output or result data**	**Access to relevant data and beneficiaries in some years**
	X	
X	X	X
X		X
X	X	X
		X
X		
X	X	X

Note: thematic ERA NETs not displayed as similar results were received
Note: "X" indicates the existence of procedure.
Source: Own data from survey to ERA NETs

TABLE 6 - Planning of impact assessments in INCO ERA- NETs

the awareness that proper documentation is important in joint funding actions was sufficiently demonstrated in the survey. In the cases where no outputs and results were collected systematically one should ask how the final reporting to the EC of the calls was organised[21].

The feasibility of future impact evaluation (e.g. lack of financing for such an activity) or the applicability of its results for follow up joint funding activities might be decisive for establishing moderate documentation and monitoring processes.

A set of questions in the survey performed by the authors explored the current evaluation efforts and the general feasibility of evaluation tasks.

Exemplary, the question whether evaluation results of previous joint funding activities could potentially shape future joint calls is answered according to the probability of a follow up. Calls do not necessarily include a follow up and therefore learning from previous funding action is not always significant. A number of the addressed joint call secretariats think that they can potentially learn more from evaluations as they currently do.

Concerning the additionality effects[22] of INCO ERA NET calls, the secretariats declare to have inquired the beneficiaries during monitoring about these effects. 60% declare to have explored behavioural additionality in the current monitoring and express that it is not possible to address that dimension more intensively in the monitoring. According to information from the survey, the control group of thematic ERA NETs do not consider behavioural additionality as relevant and express that it is not feasible to assess behavioural additionality in the future in a better way.

Concerning input additionality the picture is contrary; here thematic ERA NETs express the view that potentially the effect of additional resources triggered by the funding could be assessed fully. The horizontal ERA NETs consider the feasibility of input additionality as much smaller (34%).

Moreover, both types of ERA NETs express generally the view that net-effects could be potentially monitored, but express to have not fully addressed this dimension in the current monitoring[23].

When addressing the indicators defined in the call announcements, the thematic ERA NETs express with 100% the view that it is feasible to define indicators at funding start and to employ them for assessment at

21 The EC requires a number of outputs in the final reporting. In case the outputs at termination of funding are not collected systematically, reporting must remain incomplete.
22 One must assume that the additionality concept is not fully understood by the respondents.
23 This in our opinion - and rather speculative - could be a very opportunistic view, as some of the ERA NETs provide limited funding which could be perceived by the beneficiaries as a complementary funding source to so or so conducted research work.

the end of funding. More differentiated is the picture for the INCO ERA NETs. One explanation would be that the set of (policy driven) objectives employed lack clear indicators for measuring success of joint funding.

In a final question reference was put to information needs about RTDI evaluation issues. Evidently some experienced respondents did refuse to ascertain information needs, but a majority of 2/3 of recipients clearly declared information needs.

CHALLENGES FOR ASSESSING EFFECTS

The direct effects of joint funding for the final beneficiaries is difficult to be distinguished from external factors which are often prevailing denominators. The schedule and duration of funding, asymmetric response to calls, differing project types supporting a variety of activities with non-comparable outputs, funding asymmetry and differing efforts on partner level form an agglomerate complexity that challenges the assessment of net effects. Moreover, the lack of indicators for the internationalisation dimension is an obstacle for any assessment or evaluation. Consequently, some additional factors would not appear in pure national funding schemes, therefore measurement options are limited and request new approaches and indicators incorporating the international dimension along the policy objectives of the calls.

CONCLUSIONS - RECOMMENDATIONS

In the preparation phase of joint calls, we recommend

- to establish quantitative indicators reflecting the objectives of a call;
- to ensure that objectives of the call correspond with ex-ante selection criteria;
- a timely set up of documentation, monitoring and planned evaluation tasks;
- to take into consideration that evaluations can help justifying allocations for future funding;
- that data provision has to be a contractual obligation for beneficiaries also after completing funding and should go beyond the minimum documentation requirements set by the EC;
- that monitoring efforts also focus whether a network was newly formed with the funded project, respectively looking also on efforts of the beneficiaries to establish a lasting future cooperation;
- to carefully consider the systematic comparison of funded work conducted with the initial proposal handed in by the final beneficiaries;
- centralized monitoring efforts which can support decentralized monitoring efforts of funding partners; a minimum acceptable standard can be negotiated to avoid duplication;
- to employ knowledge from (external) evaluators or qualified staff from agencies to streamline monitoring and evaluation tasks;
- to establish learning with concluded joint funding actions (ERALEARN2, EC Annual joint funding meetings, exchange on evaluation practices in national programmes);
- to sharpen awareness on the value of monitoring and impact evaluation for justification of follow up and sustainability of funding in joint calls.

Given the broad variety of countries involved in horizontal ERA-NET calls, one must accept that also evaluation culture differs widely; therefore the potential of learning communities, the importance of transfer of appropriate practices cannot be underestimated.

The persistent pilot character of joint calls brings forward the question, whether findings, e.g. from monitoring or any evaluation, can be useful. Why should comprehensive monitoring and evaluation (summative or formative) be employed if there is no potential take up of findings? The authors cannot provide a comprehensive answer to that question, but that generally any possible learning opportunity should be employed. The chance to establish systematic documentation and monitoring in ERA-NETs should not be missed. A systematic approach can support future impact evaluations and might contribute to better justification of public spending for international co-operation.

AUTHORS

MARTIN FELIX GAJDUSEK
Centre for Social Innovation (ZSI)
E: gajdusek@zsi.at

NIKOS SIDIROPOULOS
University of Athens, Centre of Financial Studies
E: nikos.sidiropoulos@gmail.com
E: nikos_sidir@econ.uoa.gr

NEW MODES OF STAKEHOLDER INVOLVEMENT IN EX-ANTE IMPACT ASSESSMENTS

SUSANNE BÜHRER

This article intends to show benefits as well as risks when using a participatory approach in ex-ante impact assessments. The example given is the assessment of the FP7 Science-in-Society programme (SiS) where a survey-based public consultation process using basic elements of a Delphi approach was used.

THE CONTEXT

The European Union has set out for an ambitious goal that is to become a genuine innovation union, in which "research and innovation are key drivers of competitiveness, jobs, sustainable growth and social progress".[1] To this end the **Horizon 2020** strategy comes to the fore that revolves around the triad excellent science, competitive industry and a better society. Against this background, an **assessment of future options** for Science-in-Society (SiS) actions intended to provide evidence for the intervention on EU level and to develop **scenarios and options** for the implementation of Science in Society actions beyond FP7 was conducted. This study was part of the interim evaluation led by Technopolis, commissioned by DG Research and Innovation and carried out between November 2011 and October 2012[2]. Central questions of the assessment were: How to proceed with SiS actions beyond FP7? How to include the various SiS topics in **Horizon 2020**? What **policy options** are there and how to build up on what already has been established? What does the **SiS community** consider to be possible paths?

During the past years, several studies have shown a divergence between EU citizens and the goals defined for science and technology. Thus there is a strong need to improve the communication between science and society in order to motivate European citizens to become more engaged in science and to re-establish the connection between science and society.[3] Science-in-Society issues also played an important role in the formulation of the "2020 Vision for the European Research Area"[4] (ERA), the Ljubljana process (Council of the EU 2009) and the Lisbon strategy.[5] The "2020 Vision for the European Research Area" adopted in 2008 underlines that the ERA "is firmly rooted in society and responsive to its needs and ambitions in pursuit of sustainable development", confirming the orientation of the Science-in-Society Work Programme.

In the light of the rearrangements of European research funding the SiS Programme faces certain challenges. Most likely the SiS Programme is supporting the Horizon 2020 goal of creating "a better society" that addresses a number of grand challenges in health and wellbeing, food and agriculture, energy and resource efficiency etc.[6] However, the SiS Programme goes well beyond these grand challenges and builds up on a lively debate about the place and the functions of science and is complemented by democratic and participatory approaches of a deliberative democracy.[7] The SiS Programme has been evolving ever since the first debates in 2000 under the general heading of "Science, society and the citizen"[8] with the goals of bringing research closer to society, using progress responsibly and stepping up dialogue. The shift from Science and Society in FP 6 to Science in Society in FP 7 and the recent debate on **Responsible Research and Innovation (RRI)** points out to the fact that the programme gained significantly in conceptual terms in recent years. In the light of the large number of topics treated by SiS **(Ethics, Gender, Governance, Open Access, Public Engagement, Science Communication and Science Education)** it seems obvious that not one singular policy will yield the expected outcomes but a more targeted approach that takes into account the differences between the different parts of the SiS Programme is needed.

METHODOLOGICAL APPROACH

The main methodological innovation used in the course of the ex-ante impact assessment was the element of a public consultation process organized in form of a **Delphi-like European-wide online survey based on the snowball sampling technique**.

The decision to conduct a European-wide online survey was based on a variety of reasons: First of all, **social inclusiveness** and broad public engagement represent major aims of SiS and the new RRI approach

1 Science in Society (SiS) (2012): Work programme 2012. Brussels: European Commission
2 Technopolis, Fraunhofer ISI (2012): Interim evaluation & assessment of future options for Science in Society Actions. Interim Evaluation. Final report.
3 Eurobarometer (2010): Science and Technology Report. Special Eurobarometer 340/Wave 73.1.
4 http://www.consilium.europa.eu/ueDocs/cms_Data/docs/pressData/en/intm/104434.pdf
5 European Commission (2009): Challenging Futures of Science in Society. Emerging Trends and cutting-edge issues. The MASIS report.
6 http://ec.europa.eu/research/horizon2020/index_en.cfm?pg=better-society
7 European Commission 2009: Challenging Futures of Science in Society. Emerging Trends and cutting-edge issues, Report of the MASIS Expert Group.
8 European Commission 2000: Commission working document „Science, society and the citizen in Europe", SEC(2000) 1973

and should according to this also be considered within the methodological approach for the impact assessment. Secondly, workshops or focus groups are only feasible with a rather restricted number of participants where mainly the narrow community of scientists and science managers could have been involved, which may be the most vocal but certainly not the only group of stakeholders concerned. Thirdly, through the **snowball sampling technique** it is possible to reach potential hidden parts of a large and heterogeneous population, which is difficult to define at its margins.

For every SiS topic (Ethics, Gender, Governance, Open Access, Public Engagement, Science Communication and Science Education) a list of statements was developed (between seven and nine), for example in the gender field: "The total share of female researchers in the EU will be raised to 45 percent". The statements are based on a literature review, own expertise in the field and input from selected experts during a qualitative pre-test. For every statement, seven questions had to be answered concerning the following aspects: desirability, associated socio economic impacts, time frame of occurrence, political level of intervention, the most important policy measure, critical success factors, potential risks. Additionally the respondents were asked to indicate their age, sex, institutional background, country of origin and their level of expertise in the field. The online questionnaire was programmed in such a manner that the respondents were able to choose which dimension(s) they would like to treat.

The survey was launched on 14th March 2012 and was closed on 4th June 2012. The starting points were the National Contact Points for the SiS programme, members of the EPTA network, parliamentary officials (STOA, CULT etc.), National Academies of Science, research organizations like EARTO, universities, but also civil society organizations and interest groups like the European Platform of Women Scientists (EPWS), the British Educational Research Association (BERA) etc.

THE RESPONDENTS

At the initial stage we aimed for 300-500 responses. Finally we received a total number of **1097** answers. Due to the complexity and length of the questionnaire many respondents interrupted their answering. Therefore we received rather different response rates according to the SiS topics and the questions asked, for example having the highest number of answers regarding desirability and time frame of occurrence and rather low response rates regarding the risks.

Overall, 291 respondents gave information about their socio-demographic background, therefore about only one third of the people who engaged in this particular exercise can be described in more detail. Related to these 291 respondents, about half of them stem from universities and around one fifth from research institutes. Other groups like the private sector, civil society organizations and public authorities are rather seldom represented – as far as we know about their institutional background. From those respondents who indicated their socio-demographic background, 49% are female, 36% are male and 15% did not specify their sex. The majority of respondents belong to the group of people which is between 30 and 50 years old (48%), further 35% are older than 50 years, less than 30 years were only 4% and 14% did not specify their age.

The majority of those respondents who indicated their country of origin stems from Germany, followed by the Netherlands and the United Kingdom. Further countries with more than 10 respondents are Spain, France and Italy. However, smaller countries like Bulgaria, Estonia, and Lithuania etc. are also represented. Thus we can conclude that a satisfactory spread was reached

At the beginning of the questionnaire, each respondent was asked to give a self-assessment about his level of knowledge regarding the selected SiS topic. The respondents could choose between "Expert Knowledge" and "Common Knowledge". Figure 1 shows the results: Even if the majority of respondents indicates "Expert Knowledge" (59 %), the central aim of our study approach can be seen as largely achieved through the reaching of many people outside the established expert circles, as the share of respondents with "Common Knowledge" is rather high (38 %).

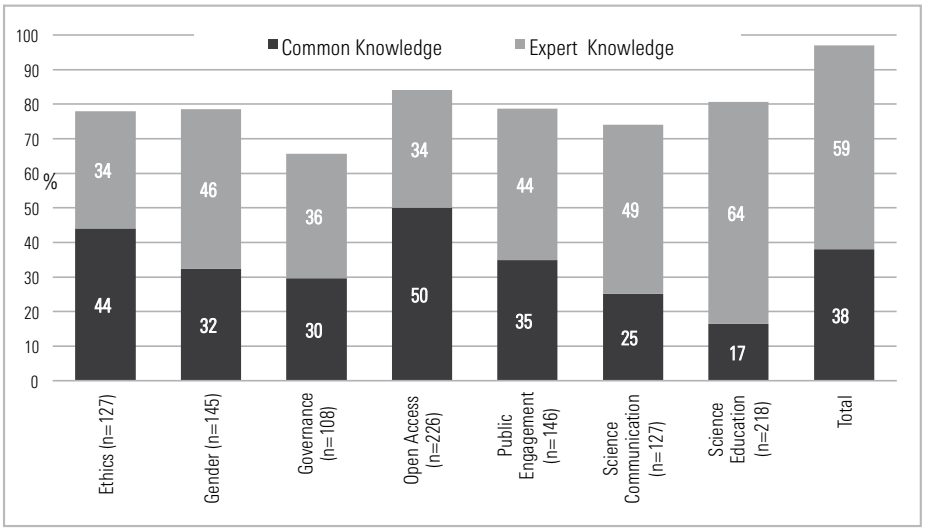


FIGURE 1 - Level of expertise based on a self-assessment

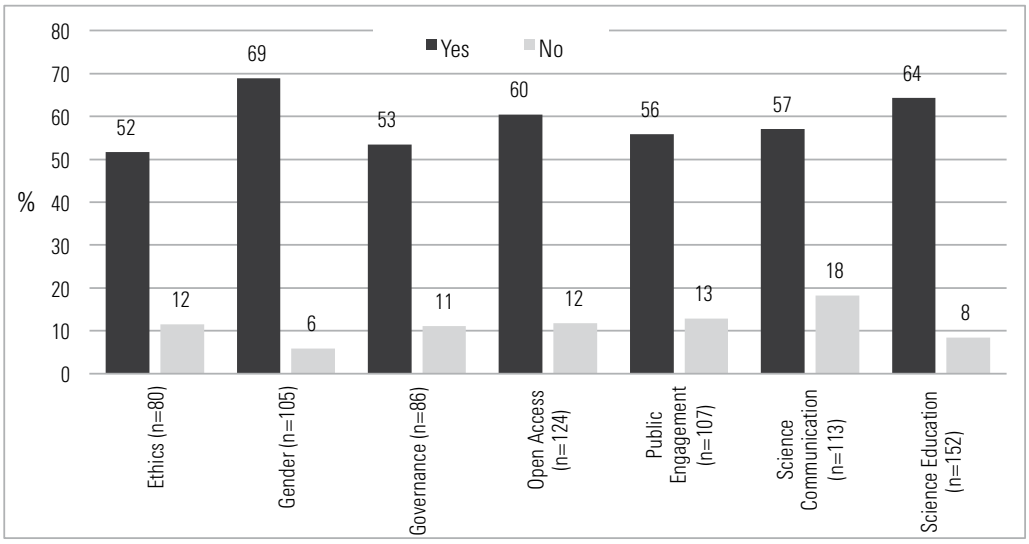


FIGURE 2 - Desirability of the statements

RESULTS

In the following we discuss the assessments of the survey respondents regarding the relevance of different thematic developments formulated within the questionnaire (the statements). The figures show the average of all answers given for the different statements in the respective topic. Overall it can be stated that much more respondents confirmed than rejected the desirability of the developments formulated in the statements (see Figure 2): At least half of the respondents assessed the different statements as desirable developments whereas a rejection occurs in maximum 18 % of all answers. When we look at the specific results, then we see that "gender" and "science education" dispose of the highest share of desirable statements.

According to the evolution and complexity of the different SiS topics as well as their broad objectives, linear cause-and-effect relationships of policy actions leading immediately to tangible outcomes are an unrealistic assumption. Additionally, from a methodological point of view, it is

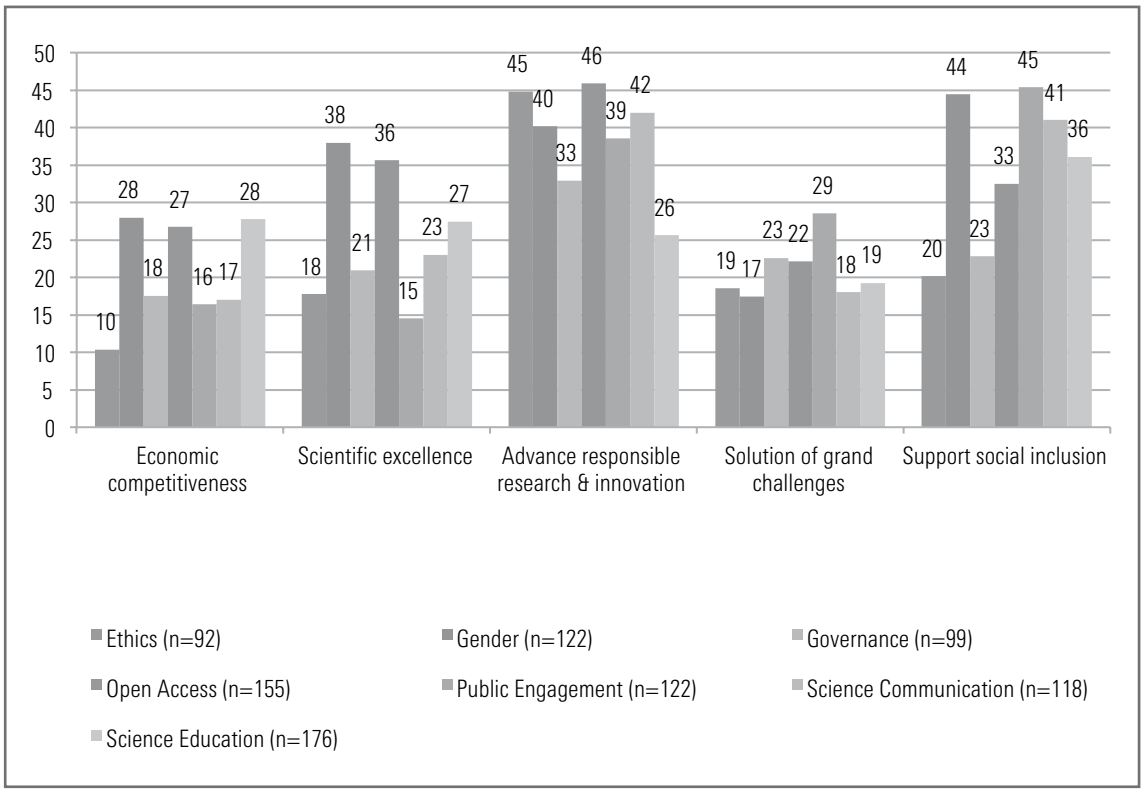


FIGURE 3 - Associated socio-economic impacts

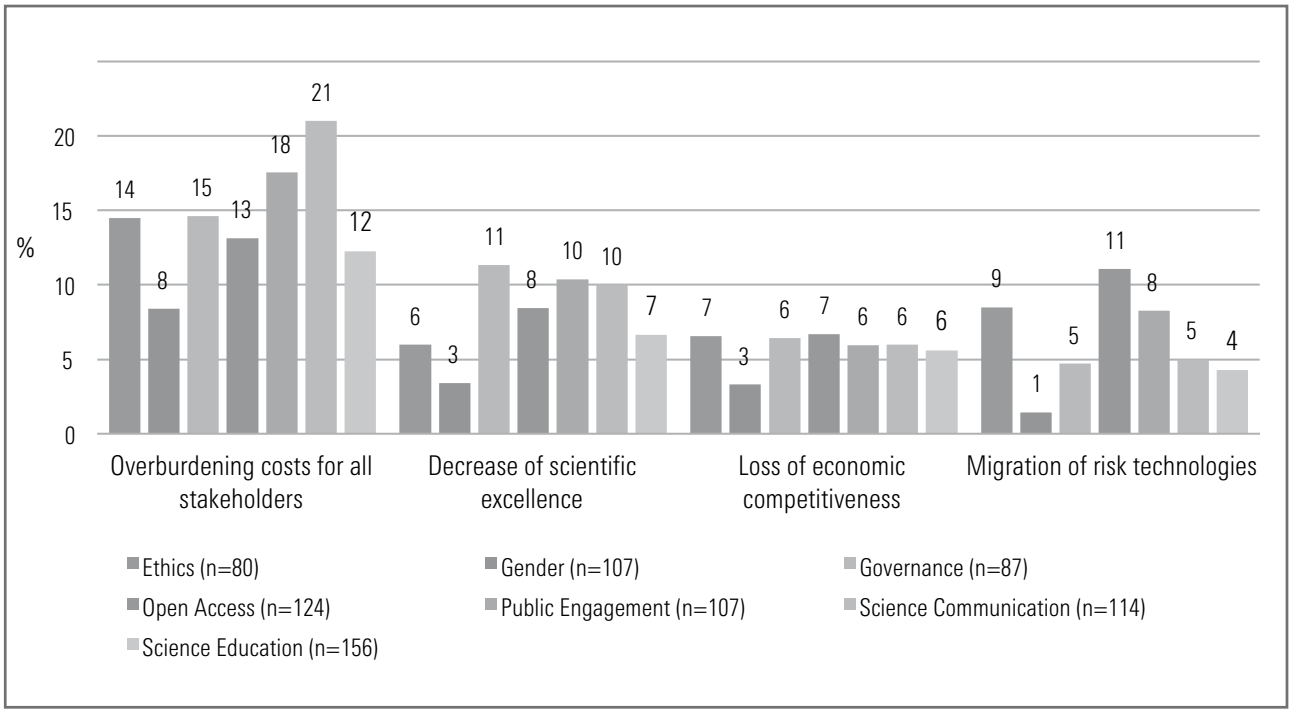


FIGURE 4 - Potential risks

rather difficult to identify measurable impacts in complex fields like the SiS ones. The first challenge is that most impacts only become visible in the long run, for example a structural and cultural change within (research) organizations as a prerequisite for a better integration of women in science. In the short run it is close to impossible to identify and measure tangible impacts. Moreover, many impacts are likely to go beyond the

intended effects and some effects may even be unintended. Therefore, a narrow cost-benefit analysis is inappropriate, especially at such an early stage.

However, the survey suggests several ways through which SiS may potentially impact on the crucial challenges of European science and innovation policy. During the survey, the strongest effects are expected with regard to the advancement of Responsible Research and Innovation (RRI) whereas particularly ethics, open access and science communication are seen as main influencing factors.

On the second place we find the impact on social inclusion where the highest shares can be found with public engagement and gender. Scientific excellence, the impact dimension which follows at the third place, is mainly affected by three factors, namely gender, open access but also science education (see Figure 3).

Finally, according to the survey respondents, impacts on the economic competitiveness are mainly associated with the human resources-related topics like gender and science education but also by open access.

With regards to the potential risks, the survey respondents indicate much fewer risks than (positive) impacts: the maximum percentage of risks mentioned is 21 % whereas the maximum percentage for a positive impact is 46 %. However, if risks are associated with the different statements, they typically refer to the challenge of overburdening costs. This result applies to all dimensions investigated, with gender showing the lowest cost risks (and other risks too) (see Figure 4).

The risk of a decrease of scientific excellence, which is the second important risk dimension, is mainly associated with governance, public engagement and science communication. This result can be interpreted as the concern that (too) much time is needed for matters which are not directly related to the conduct of actual research in a narrow sense and afterwards this time is missing for the basic tasks. The high scores for the governance dimension can be related to the fact that the public consultation survey referred at this point very much to the relationship between the business sector and other parts of society which are seen as important contributors in the decision-making process on future research priorities.

These results confirm a remarkable uncertainty among European scientists or at least those researchers who participated in the survey, how to cope with the challenges regarding a better integration of societal (and industry-related) issues in science and prove at the same time that it is not yet fully understood that the consideration of societal needs have the potential to enlarge and improve scientific discoveries

CONCLUSION

The public consultation survey delivered **substantial new evidence** on questions regarding the different SiS topics and future options of the SiS programme. As intended, a large and heterogeneous population could be reached with the survey. In addition to the experts, which made up 59% of the respondents, 38% of the participants indicated only "**common knowledge**" in the respective area.

In the end, the experts and lay people involved in the survey gave a very **positive feedback** on this particular way to organize a public consultation process. Through the complex set of questions and the rather high number of respondents, a substantial set of relevant findings could be generated which go far beyond the narrow project context.

A certain weakness of the snowball-sampling approach consists in the uncertainty about the respondents: typically a questionnaire design foresees that the questions regarding the socio-demographic background are not posed before the latest content-related issue. But if – as in our case – only rather few respondents reach this later part, crucial information about the background of the respondents is missing.

Additionally, as the methodological approach was completely different compared with conventional ex-ante approaches, it was rather difficult (if not impossible) to convince the client of the strengths and advantages of the methods used and the robustness of the empirical results.

AUTHOR

SUSANNE BÜHRER
Fraunhofer Institute for Systems and Innovation Research
E: s.buehrer@isi.fraunhofer.de

EVALUATION AS THE CONSTRUCTION OF POLICY NARRATIVES

ERICH PREM

INTRODUCTION

It has happened largely unnoticed, but we live in a world of evaluation. From schools to restaurants, from transportation to accommodation, everything undergoes evaluation today. Much of this trend is facilitated by the internet and its possibilities for peer and customer review. But we also experience a massive trend towards professional evaluation and public policies are no exception to this development. In fact, professional, third-party expert policy evaluation has now become state-of-the-art in policy making for many countries and organisations. It has also found its way into regulation and good public management practices. While it is true that there is the occasional excess of programme and policy evaluation, this overall tendency is certainly in line with a need for efficient and effective public policies.

The massive trend to professional evaluation also poses the question as to what the role of the professional evaluator is in the overall public policy process. The evaluator can be regarded as a referee, ensuring the proper application of the rules of the political "game". From another perspective, s/he may be seen as the wise guy or *guru*, an external authority providing a source of lore, political acumen, and sagacity. Thirdly and perhaps more in line with the self-image of evaluators, s/he can be regarded as an expert or consultant – in this case of the how-to of the evaluation business. All these different role models position the evaluator as clearly distinct and apart from the subject matter of evaluation. In these views, the evaluator is not a direct participant in the process of policy creation; he or she remains an external advisor. However, in contrast to these other roles, the evaluator may also be viewed as a contributor to policy narratives, i.e. as a creator of the political storyline and framework ultimately justifying political choices. This is the role which we investigate in more detail in the remainder of this paper.

EVALUATION: ASSESSMENT OR APPRECIATION?

Evaluation has often been defined as the "systematic and objective assessment" of a planned, on-going or completed project or programme, its design, implementation and results. Evaluation in this view aims to determine the relevance and fulfilment of objectives and the efficiency, effectiveness, impact and sustainability of the programme or project under evaluation.[1] This definition, also used by the OECD, creates an objectivist picture of evaluation. It emphasizes the instrumental and therefore the technical nature of public programmes. In such a view, evaluation serves to examine the function of the programme assuming a rational framework in which programme objectives and instruments have been clearly described and in which the instruments were constructed so that their anticipated consequences serve to realize those objectives. In this view, programmes are primarily regarded as rational means to reach clearly defined aims.

In a rather different explanation (Scriven, 1999), "evaluation is the systematic investigation of the merit, worth or significance of an object." In this meaning the process of evaluation assigns value to a program's efforts by addressing three inter-related domains: merit (or quality), worth (or value, i.e. cost-effectiveness) and significance (or importance). This definition of evaluation stems from a psychological background and its immediate connection to the field of programme evaluation may not be obvious. However, it is argued here that evaluation necessarily carries such an element of value creation. This is particularly true for science, technology and innovation (STI) evaluation – and even more so in its practice rather than its theory. This aspect of value creation is significant regarding the nature of evaluation as it constitutes a step in the creation (or co-creation to be precise) of policy narratives, as we will discuss below.

THE INSTRUMENTAL VIEW OF RTDI POLICY MAKING

Public research, innovation, and technology programmes are instruments in a political context. They serve to pursue goals which have been recognized as important by political decision-makers. In this sense, programmes are embedded in a framework of policy narratives justifying the public intervention as useful and goal-driven. Indeed, often a first step in programme evaluation is the clarification of the underlying intervention logic, for example using so-called logic charts. This approach follows the logic model developed in the 1970s by Caro Weiss and Joseph Wholey which has been refined many times since then. Such a logic model also subscribes to a rationalist, objectivist and instrumental view of STI programme design and evaluation. The categorical distinction between inputs, outputs, outcomes and impacts has become a classic in STI programme design and evaluation.

The logic model approach typically constitutes one of the first steps in the evaluator's task. On the other hand, in modern STI programme design, it is typically assumed that the underlying rationale for any pro-

1 OECD DAC Glossary

gramme intervention is such a logic model according to which the programme was first designed. In this sense, logic charts of a programme make the underlying design explicit and a careful analysis at this stage may already reveal differences between the design or intent and implementation of an STI programme, i.e. its actually achieved or expected outputs, outcomes, and impacts. Of course, expected outcomes and impacts need to be interpreted and understood within a larger context of policy making. Programme evaluation can also be regarded as a process of validating the intervention with respect to the overarching policy objectives. In this understanding, evaluation often implicitly tests the plausibility of the policy narrative in which any public programme is typically embedded. A programme failing to deliver the intended outcomes and impact may very well question the underlying policy rationale – rather than just the functionality of the applied instruments. This to some extent may blur the boundaries of the subject of evaluation: it could be that the instruments are not fit to the task at hand, but it may also be that the policy framework and narrative are unveiled as unfit and erroneous. This aspect of evaluation is hardly surprising for the programme evaluation professional, still it is not typically made explicit in programme evaluation. Rather, the embedding policy narratives are often taken as a given in the practice of evaluation.

EVALUATION AND THE CREATION OF POLICY NARRATIVES

Most importantly, evaluation often constitutes an important part in creating or at least co-creating (i.e. contributing to) policy narratives or components in such narratives. There are several reasons why evaluators become creators of policy narratives in the practice of programme evaluation. Obviously, there are cases where programmes were not completely specified in full detail with respect to the logic of intervention. Delivering an explicit description of a programme logic in such cases will almost automatically generate options for renewed versions of the policy narrative supporting the intervention. Although this may not be considered the optimum case, there are often practical reasons for this situation including for example the pressure to meet deadlines for starting a programme etc. But even when there is an adequate logic describing the intervention, evaluation will typically create additional components of the policy narrative. Evaluation often adds to the various dimensions of the policy context by finding and defining new objectives, causal or statistical relations, or simply "anecdotal evidence". Such additions, examples, components etc. can easily be taken as arguments of policy makers in their narratives and they may also lead to new, more refined logic models.

It is worth pointing out that the identification of the programme logic is a creative act and already part of the interpretation process. Even the description of the programme objectives typically is not just a case of copying from the programme document. Evaluators typically perform interviews or study preparatory documents and workshop reports to identify the real programme objectives. Even these innocent looking acts constitute elements in the potential construction of the policy narrative.

This brings us back to the second, more psychological definition of a discipline of evaluation devoted to a systematic determination of merit, worth, or significance (Scriven, 1999). Apart from the establishment of factual premises (such as performance or impact) the evaluation exercise includes the identification of value premises such as those both embedded in and expressed in policy narratives. Today, most programme evaluators would not consciously regard their work as creating policy narratives (perhaps with the notable exception of ex ante evaluation).

This aspect of the role of the programme evaluator has only found little attention in the evaluation literature. In practice, however, evaluation as co-creation of policy narratives is hardly avoidable. It is deeply rooted in the nature of any valuation, i.e. the determination of merit and worth. Once statements about merits are created they are easily turned into value statements in the context of political justifications.

In a broader view, not all the actions and decisions of policy makers are easily described in an entirely rational framework. This does not make them irrational, but they can often more easily be understood with reference to emotions, opinions, perceptions rather than rationality, facts and figures. The evaluator only reflects this aspect of policy making when making value judgements. Perhaps paradoxically for all the intended rationality, the evaluator also necessarily delivers valuation and appreciation.

CONCLUSIONS & REFERENCES

From a philosophical perspective, our line of thinking may be considered post-positivist (cf. Gottweis 2003) rather than just standing in an analytic tradition. The new role of the evaluator does neither arise from a mere analysis of the meaning of "value" nor is it limited to an interpretative or hermeneutic account of what is involved in preparing evaluations. The new view follows from a focus on the social relations among the actors involved in evaluation and policy making. The view of the evaluator as a participant in the creation of policy narratives arises directly from questioning the purely objectivist self-image of the evaluation action and from unveiling the often implicit, but creative act of valuation.

Our analysis of evaluation suggests that programme evaluators in principle cannot avoid their participation in policy creation. This is a consequence of what we mean by "value" – it follows from the *nature* of evaluation. The important question then is how to deal with this consequence in the practice of programme evaluation. Certainly, it is essential to accept the potential contribution of evaluation to the surrounding policy narrative as a potential outcome of the evaluation endeavour. Even more so, it may be worthwhile for the evaluator to actively embrace the role as a creator of policy narratives and to make it as explicit as possible.

Scriven, Michael (1999): The nature of evaluation, ERIC/AE Digest Series EDO-TM-99-06, University of Maryland, College Park, US. Also available from http://ericae.net/digests/tm9906.pdf

Gottweis, Herbert (2003): Post-positivistische Zugänge in der Policy-Forschung. Politik als Lernprozess. Springer.

AUTHOR

ERICH PREM
eutema GmbH
E: prem@eutema.com

WIE WIRKSAM SIND INNOVATIONSFÖRDERMASSNAHMEN IN DER SCHWEIZ?

FRANZ BARJAK

Der vorliegende Beitrag fasst eine Sekundäranalyse von 16 Evaluationen und Studien der Innovationspolitik der Schweiz aus dem Zeitraum 1997-2012 zusammen (Barjak, 2013). Neben einem begleitenden Tracking und einer nachgelagerten Leistungsmessung sind periodische Synthesen einzelner Evaluationen sehr sinnvoll. Sie tragen zu einem differenzierteren Verständnis von Innovationsvorgängen in Unternehmen bei und beleuchten die Reaktionen der Akteure auf die Massnahmen zur Innovationsförderung (und ihre Evaluation). Erst die Zusammenschau von Einzelevaluationen ermöglicht es, die Wirksamkeit, Stärken und Schwächen der Innovationsförderung auf Systemebene und über einzelne Akteure und Interventionen hinweg zu diskutieren (Arnold, 2004; Edler, Ebersberger & Lo, 2008; Magro & Wilson, 2013). Sie hilft dabei, den aktuellen Kenntnisstand zur Innovationsförderung in einer Volkswirtschaft zu resümieren. Die einbezogenen 16 Innovationsfördermassnahmen der Schweiz lassen sich zu vier Typen zusammenfassen:

- **Technologieorientierte Programme** (CIM, Microswiss, Soft[net], TOP NANO 21, Medtech, Diffusion energieeffizienter Technologien). Bei diesen Programmen geht es vor allem darum, die Anwendungs- und/oder Produktionskompetenz der Unternehmen in bestimmten Technologiebereichen (Computer-Integrated Manufacturing, Microtechnologien, Software, Nanotechnologie) zu steigern. Die Programme sind vor allem diffusionsorientiert, ausgenommen die Medtech-Initiative, bei der die Steigerung der Innovationskompetenz in der Medtech-Branche im Zentrum stand. Medtech und TOP NANO 21 zielten ausserdem auf den Kompetenzaufbau im Bildungs- und Forschungssektor ab, wobei TOP NANO 21 eine starke Wissenschafts- und Grundlagenkomponente enthielt. Zielgruppen der Förderung sind jeweils Bildungs- und Forschungseinrichtungen sowie Unternehmen – beim Programm zur Steigerung der Energieeffizienz nur letztere. Bei frühen Programmen (CIM, Microswiss) wurden eigene Zentren aufgebaut, die spezifische Dienstleistungsangebote anbieten sollten. Die Programme ab dem Jahr 2000 sehen nur noch den Aufbau virtueller Zentren durch Vernetzung bestehender Institutionen und Unternehmen vor. Im Kern bleibt das Dienstleistungsangebot jedoch ähnlich und umfasst Aus- und Weiterbildung, Forschung und Entwicklung sowie Technologietransfer. Im Rahmen von Soft[net] und Top Nano 21 werden zudem noch Gründungsaktivitäten unterstützt.

- **Reguläre KTI-Projektförderung:** Drei der einbezogenen Arbeiten beschäftigten sich hauptsächlich mit der regulären Projektförderung der Kommission für Technologie und Innovation (KTI), der Innovationsförderagentur des Bundes. Diese Förderung war zudem auch der zentrale Inhalt einer Studie zum Kompetenzaufbau an Fachhochschulen. Die Ziele der KTI-Förderung sind wirtschaftliche Innovationsprozesse durch projekt- und programmspezifische Förderung anwendungsorientierter FuE zu unterstützen und den Auf- und Ausbau einer wettbewerbsfähigen anwendungsorientierten FuE an den Hochschulen zu fördern (Grunt, et al., 2003, S. 26-27). Damit sind auch die beiden Hauptzielgruppen der KTI-Förderung beschrieben. Die reguläre Förderung umfasst im Kern die finanzielle Unterstützung gemeinschaftlicher FuE-Projekte von nicht-gewinnorientierten Forschungs- und Bildungsstätten und Wirtschaftspartnern. Dabei müssen die Wirtschaftspartner mindestens 50% des Projektvolumens beisteuern und die KTI-Zuschüsse werden nur an den Forschungspartner ausgerichtet.

- **Wissens- und Technologietransfer (WTT) Förderung:** Die WTT-Konsortien stellten eine diffusionsorientierte Massnahme dar und waren von 2005 bis 2012 aktiv. Sie unterstützten vor allem die Kommunikation des Transferbedarfs von Unternehmen an Hochschulen (Pull-Prozess), aber auch den Transfer von Hochschulen zu Unternehmen (Push-Prozess) sowie die regionale bzw. (in einem Fall) nationale Koordination des WTT. Ihre Coaches und Experten boten insbesondere Beratung für KMU zu spezifischen Fragen des WTT an. Ferner vermittelten sie Partner und Informationen. Die regionalen WTT-Konsortien wurden zum 01.01.2013 von Nationalen Thematischen Netzwerken (NTN) abgelöst.

- **Start-up Förderung** (KTI Start-Up Label, Venturelab): Die Start-up Förderung soll Unternehmensgründungen erleichtern und die Erfolgsrate neu gegründeter Unternehmen verbessern. Sie ist ebenfalls eine Form der WTT-Unterstützung. Während Venturelab mit der Kompetenzvermittlung und Sensibilisierung im Bereich Entrepreneurship vor allem potenzielle Gründer anspricht, bietet KTI Start-up Dienstleistungen für Jungunternehmerinnen und -unternehmer an. Wichtig sind in beiden Programmen personenbezogene Qualifikationsmassnahmen (Workshops, Seminare, Kurse) und die Vernetzung der Gründer und Gründerinnen untereinander, aber auch die Vernetzung mit Personen, die den Gründungsprozess unterstützen, wie etwa Business Angels, Venture Capitalists oder Mentoren.

Einen Überblick über die unterschiedlichen Fördermaßnahmen bietet Tabelle 1.

	Evaluationszeitraum	Volumen der öffentlichen Förderung in diesem Zeitraum in SFr.				Gesamtvolumen der Projekte in SFr.	
		Gesamt (in Mio.)	Anzahl Projekte	pro Projekt	pro Jahr (in Mio.)	Gesamt (in Mio.)	pro Projekt
1. **CIM-Aktionsprogramm FH-isi**	1990-96	102			14.6		
2. **CIM-Aktionsprogramm KOF**	1990-96						
3. **Microswiss**	1991-96	110			18.3		
	1991-96	65.1b	318b	205'000b	10.9	115.9c	464'000c
4. **Evaluation SNF/KTI**	2000-03	320			80		
	1995-2000	370	1'700	218'000	61.7	1'040	612'000
5. **Softnet**	2000-03	30	151	199'000	7.5		
6. **TOP NANO 21**	2000-03	72	260	277'000	18	109	419'000
7. **KTI-Projektförderung**	2000-02	120.9	634	191'000	40.3		
8. **Medtech-Initiative**	1998-2003	36	134	269'000	6	90.7	677'000
9. **Dissertation zur KTI**							
10. **Angewandte FuE an FH**	1998-2007	215			21.5		
	1998-2004	141	772	183'000	20.1	347	449'500
11. **Konzeptevaluierung WTT-Initiative**							
12. **Start-up Label 2006/07**	1996-2005		153				
13. **Venturelab Entrepreneurship, Education & Training**							
14. **WTT-Initiative**	2005-10	23.8			4		
15. **Start-up Label 2011**	1996-2009		243				
16. **Diffusion EET**	2008	106			106		

TABELLE 1 - Finanzielles Volumen der Fördermassnahmen[a].

a. Angaben auf der Basis der evaluierten Studien. Sie wurden nicht mit anderen Quellen verglichen und auch nicht auf Konsistenz geprüft. Leere Zellen bedeuten „keine Angabe" in der einbezogenen Literatur.
b. Nur Industrieprojekte.
c. Ohne KTI-Sonderkredit.

SCHWEIZER EVALUATIONSPRAXIS IST VERGLEICHBAR ZU DER IM ÜBRIGEN EUROPA

Die Meta-Evaluation im Rahmen der Untersuchung zeigte, dass die Praxis der Evaluation von Innovationspolitik in der Schweiz hinsichtlich Methoden und Ergebnissen mit der Praxis in anderen europäischen Ländern vergleichbar ist (vgl. dazu Edler, Dinges & Gök, 2012). Allerdings gibt es auch Verbesserungsmöglichkeiten im Hinblick auf 1) die Beschreibung von Evaluationskonzeption und Methoden, 2) die Auswahl von Kontrollgruppen, 3) die Kombination quantitativer und qualitativer Methoden und 4) die Berücksichtigung aller Stakeholder einer innovationspolitischen Intervention.

KASTEN 1: SEKUNDÄRANALYSEN VON EVALUATIONEN

Ziel von Sekundäranalysen ist die Ermittlung verlässlicherer Ergebnisse durch die Berücksichtigung der Methoden und des Kontexts von Studien und Synthese ihrer Inhalte. Drei Arten von Sekundäranalysen leisten spezifische Beiträge (Widmer, 1996; Denyer & Tranfield, 2006):

- **Meta-Evaluationen** sind Studien, die Evaluationsstudien bewerten und anhand Kriterien wie Nützlichkeit, Anwendbarkeit, Korrektheit, Genauigkeit etc. evaluieren.
- **Für Meta-Analysen** gibt es verschiedene Definitionen, in aller Regel wird aber damit eine statistische Synthese quantitativer und systematisch ausgewählter Analysen bezeichnet.
- **Evaluationssynthesen** geben einen Überblick über die Ergebnisse quantitativer und qualitativer Studien und nutzen in der Regel keine statistischen Verfahren.

DIE INNOVATIONSFÖRDERUNG DER SCHWEIZ IST GEMÄSS BISHERIGEN EVALUATIONEN KONSISTENT, WIRD WEITGEHEND EFFIZIENT UMGESETZT UND ERREICHT IHRE ZIELE

Die ausführliche Evaluationssynthese im Rahmen dieser Arbeit lieferte Erkenntnisse zu Konsistenz, Effizienz, Zielerreichung und Wirkungen der Innovationsförderung.

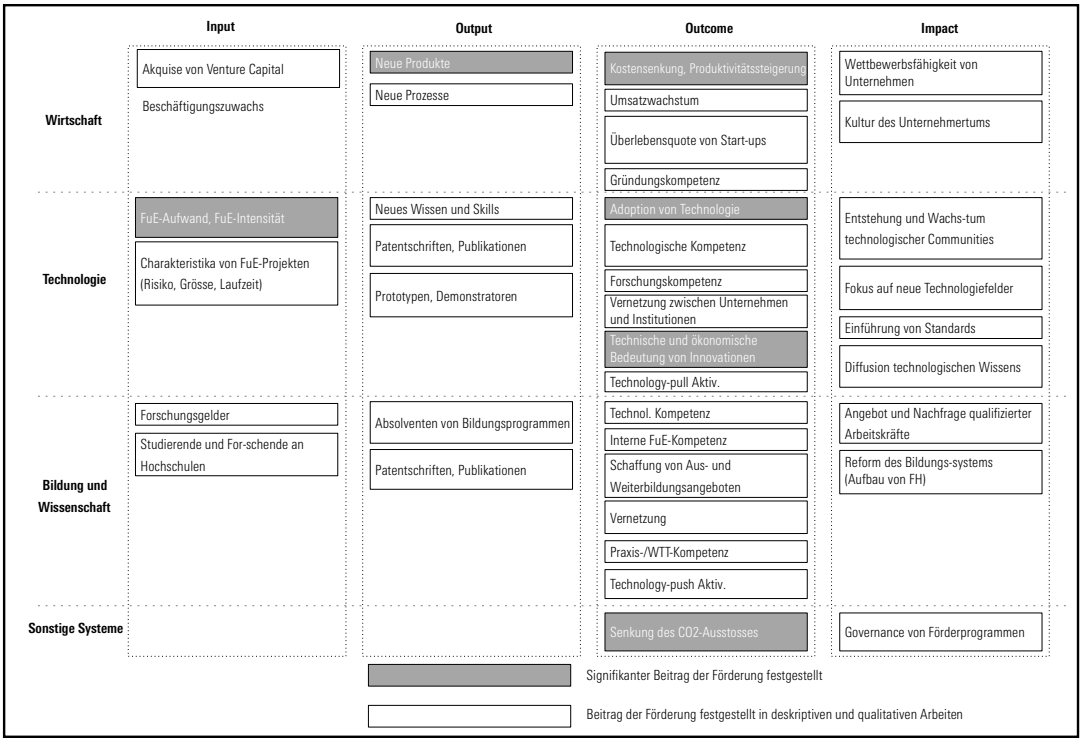
Die Evaluationen beurteilen die Konsistenz oder Eignung der innovationspolitischen Fördermassnahmen zur Lösung der vorab identifizierten Probleme und Kohärenz zu anderen Massnahmen und Institutionen überwiegend als gut. Die meisten Probleme wurden bei den Ausführungsbestimmungen gesehen, die Förderziele und vorgesehenen Zeitrahmen der Förderung nicht immer in Einklang brachten.

Auch für die Implementation wurden nahezu durchwegs gute Noten verteilt, wobei grössere Fördermassnahmen (TOP NANO 21, Medtech) tendenziell etwas besser als kleine (Softnet, angewandte FuE an FH) abgeschnitten haben. Die Frage, warum dies so ist, kann auf der Basis der Sekundäranalyse nicht beantwortet werden. Einzig der Kommunikation und Koordination zwischen den im Rahmen der Förderung geschaffenen Institutionen wurde bei einigen Massnahmen ein schlechtes Zeugnis ausgestellt. Dies führte sowohl zu Abstimmungsschwierigkeiten in der Umsetzung von Massnahmen als auch zum Ausbleiben von Spezialisierungs- und Lerneffekten.

Die Innovationsförderung erreicht gemäss dem Urteil der Evaluationen in den meisten Fällen ihre technologischen Ziele. Abstriche gibt es bei den davon abgeleiteten wirtschaftlichen Zielen: eine Markteinführung von Innovationen, Kommerzialisierung von Projektergebnissen, Stärkung einer Branche durch neue Produkte und Gründungen konnte nicht immer wie vorab formuliert, und eigentlich intendiert, realisiert werden.

VIELFACHE EFFEKTE DER INNOVATIONSFÖRDERUNG FESTGESTELLT

Die Förderwirkungen sind in der Abbildung 1 zusammengefasst.[1] Die Konstruktion der Innovationsförderung stellt weitestgehend sicher, dass durch die Förderung zusätzliche FuE-Ausgaben getätigt werden. Die Evaluationsstudien legen auch nahe, dass geförderte Projekte grösser sind und schneller durchgeführt werden können. Allerdings zeigt sich, dass die Outputs der Förderung (siehe Kasten 2) nur selten quantifiziert werden konnten. Die dafür erforderlichen Daten fehlen vielfach. Die Outcomes der Förderung, wie z.B. Kostensenkungen, Umsatzwachstum, Erweiterung der technologischen Kompetenzen, oder Vernetzung wurden in den Evaluationsstudien deutlich besser abgebildet und zum Teil



Anmkg: Der Beschäftigungszuwachs bezieht sich insbesondere auf Zuwachs von F&E-Personal.

ABBILDUNG 1 - Übersicht über die festgestellten Auswirkungen der Schweizer Innovationsförderung

1 Für die Abbildung wurde eine Vorlage auf der Basis ähnlicher Untersuchungen in Finnland adaptiert (vgl. Hyvärinen, 2011).

in Arbeiten mit einer statistischen Fundierung bestätigt. Die Ermittlung von Impacts der Förderung auf Wirtschaft und Gesellschaft insgesamt scheitert bislang aus mehreren Gründen: dem kurzen Zeithorizont der Evaluationen, dem vergleichsweise geringen Fördervolumen und den umfassenden Daten und konzeptionellen Modellen, die für eine solche Analyse notwendig wären, bislang aber nicht vorliegen.

Auswirkungen der Förderung auf das Verhalten der Geförderten, also etwa eine höhere Forschungskompetenz, Gründungsneigung oder Bereitschaft zu WTT-Projekten wurden in den Evaluationen zwar ebenfalls vielfach vermutet; sie sind aber bislang noch kaum durch messbare Kriterien und Daten unterlegt.

EMPFEHLUNGEN ZUR EVALUATION INNOVATIONSPOLITISCHER INTERVENTIONEN

Die betrachteten Studien geben zunächst eine Reihe von Anhaltspunkten dafür, dass die Innovationspolitik der Eidgenossenschaft grundsätzlich ihre Ziele erfüllt und positive Auswirkungen auf die Innovationstätigkeit in der Schweizer Wissenschaft und Wirtschaft hat. Im Hinblick auf eine weitere Verbesserung der Politik und Erweiterung des Wissens zu ihren Wirkungen wären folgende Massnahmen zielführend:

1. Die Kommunikation während der und über die Förderung sollte ausgeweitet werden, um die jeweiligen Fördermöglichkeiten und Bestimmungen den Zielgruppen noch besser zu vermitteln.
2. Im Rahmen einer übergreifenden Evaluationskonzeption sollte festgelegt werden, wann und mit welchen Zielen und Methoden welche Art von Evaluation durchgeführt werden muss. Damit liessen sich die Verbindlichkeit und Relevanz von Evaluationen steigern.
3. Technologische Wirkungen und Auswirkungen der Innovationspolitik auf Bildung und Wissenschaft sollten kontinuierlich identifiziert werden, wofür etwa ein Indikatorsystem zu etablieren wäre.
4. Umfassendere Anforderungen an Evaluationen müssen mit einer entsprechenden Mittelzuweisung für Evaluationen einhergehen.

KASTEN 2: DIMENSIONEN VON WIRKUNGSANALYSEN

Wirkungsanalysen von innovationspolitischen Interventionen differenzieren vielfach zwischen drei Dimensionen (vgl. Falk, 2007; Good, 2005):

1. dem gesellschaftlichen Subsystem, in dem eine Wirkung auftritt, z.B. Wirtschaft oder Wissenschaft;
2. wann und bei wem Wirkungen auftreten, wobei Outputs (kurzfristig, direkt Beteiligte), Outcomes (mittel- und langfristig, direkt Beteiligte) und Impacts (mittel- und langfristig, Nicht-Beteiligte) unterschieden werden;
3. Intervention sollten zusätzliche Leistungen erzielen und nicht etwa Leistungen anderer verdrängen (z.B. über Mitnahmeeffekte), dadurch dass sie zusätzliche Inputs mobilisieren (**input additionality**), zusätzliche Outputs generieren (**output additionality**), oder das Verhalten der Geförderten beeinflussen (**behavioural additionality**).

BIBLIOGRAPHIE

Arnold, E. (2004). Evaluating research and innovation policy: a systems world needs systems evaluations. Research Evaluation, 13(1), 3-17.

Barjak, F. (2013). Wirkungen innovationspolitischer Fördermassnahmen in der Schweiz. Im Auftrag des Bundesamtes für Berufsbildung und Technologie BBT. http://www.sbfi.admin.ch/dokumentation/00335/01740/index.html?lang=de&download=NHzLpZeg7t,lnp6I0NTU042l2Z6ln1acy4Zn4Z2qZpnO2Yuq2Z6gpJCEdX1,e2ym162epYbg2c_JjKbNoKSn6A--

Denyer, D., & Tranfield, D. (2006). Using qualitative research synthesis to build an actionable knowledge base. Management Decision, 44(2), 213 - 227.

Edler, J., Berger, M., Dinges, M., & Gök, A. (2012). The Practice of Evaluation in Innovation Policy in Europe. Research Evaluation, 21(3), 167-182.

Edler, J., Ebersberger, B., & Lo, V. (2008). Improving policy understanding by means of secondary analyses of policy evaluation. Research Evaluation, 17(3), 175-186.

Falk, R. (2007). Measuring the effects of public support schemes on firms' innovation activities: Survey evidence from Austria. Research Policy, 36(5), 665-679.

Good, B. (2005). Technologie zwischen Markt und Staat: Die Kommission für Technologie und Innovation und die Wirksamkeit ihrer Förderung. Zürich & Chur: Verlag Rüegger.

Grunt, M., Reuter, A., & Heinzelmann, E. (2003). Evaluation der Kommission für Technologie und Innovation. Bericht „Selbstevaluation" (1 ed.). Bern: Bundesamt für Berufsbildung und Technologie (BBT).

Hyvärinen, J. (2011). TEKES impact goals, logic model and evaluation of socio-economic effects. Research Evaluation, 20(4), 313-323.

Magro, E., & Wilson, J. R. (2013). Complex innovation policy systems: Towards an evaluation mix. Research Policy, 42(9), 1647-1656.

Widmer, T. (1996). Meta-Evaluation - Kriterien zur Bewertung von Evaluationen. Bern: Verlag Paul Haupt.

AUTOR

FRANZ BARJAK
Hochschule für Wirtschaft
Fachhochschule Nordwestschweiz (FHNW)
E: franz.barjak@fhnw.ch

PORTFOLIO EVALUATION: A CASE STUDY TO ILLUSTRATE EVALUATION CHALLENGES

CHRISTIANE KERLEN, JAN WESSELS, VOLKER WIEDEMER

INTRODUCTION

Portfolios are of increasing interest in programme planning and implementation but even more important in monitoring, optimising and steering of programme families that already exist. Portfolios are understood as a group of parallel programmes or individual measures that are directed towards the same target group of participants and which may be evaluated jointly (Fischl/Kulicke/Wessels 2013). The programmes or measures within a portfolio should be considered as partially independent (they could also work as standalone programmes or measures), but also interdependent (with e.g. synergies or distorting effects). Portfolios can be made up of regions, certain fields of technology or innovation, or agencies with the same target group. Problems related to the definition and grouping of these portfolios concern steering, legitimising, competition, allocation (Jörg in Fischl/Kulicke/Wessels 2013). The increasing orientation of innovation policy towards a challenge oriented policy and systemic change leads to increased use of programme portfolios, which try to address different aspects of the innovation system in an integrated way. This means that portfolios become more frequent in innovation policy, even if integrated evaluations of such portfolios remain rather rare up to now.

EVALUATING PROGRAMME PORTFOLIOS

There are a variety of dimensions that add to the complexity of a portfolio[1] and therefore raise a number of problems in terms of comparability and overall assessment. These include the following:

- Number of instruments: Various different instruments might be grouped into one portfolio that have different mechanisms of impact related to them. The systemic approach of actual innovation policies is implemented by portfolios of instruments which address very different aspects of the respective system like cooperation of stakeholders, legal and fiscal framework conditions, aspects of learning and education and so on.
- Diversity of stakeholders: Grouping different programmes into one portfolio means that the number and diversity of stakeholders increases. These include different subgroups within the target groups of the programmes.
- Variety of goals/overlapping goal systems: If a portfolio consists of different programmes or measures there will probably be a variety of goals. In general those goals should match which each other and should also be the reason for grouping these programmes into one portfolio. But there may still be diverging priorities or even opposite sub-goals. This means that there may be effects going into different directions, which pose a challenge for assessing if goals have been met.
- Institutional frameworks: There may be different institutions involved in one portfolio. This might be true for programme agencies as well as different governmental bodies as programme owners. Challenge oriented policy strategies tend to include several ministries with interrelated, but also competing programmes and measures.
- Industries/application fields: If the field of technology or innovation is not the reason for establishing a certain portfolio, it is likely that there will be a variety of different fields of innovation or different industries resembled in one portfolio. Converging technologies and systemic approaches bring together very different technologies, e.g. for the case of electro mobility, the traditional industry of car manufacturers must now interact with energy suppliers etc.

The increasing complexity and interconnectedness of programmes and instruments on one side leads to increasingly complex evaluation designs. Evaluations have to fulfil more tasks and requirements at the same time which are interconnected and complex in itself. This results in complex requirements regarding the design of an evaluation, the role of the evaluator, and the methods employed.

To evaluate a portfolio – and this point we would like to illustrate in this article – makes an evaluation multidimensional and raises its complexity. Among the dimensions are:

- Evaluation function: Whereas the legitimising function is prevalent in an ex-post perspective, the learning function becomes predominant in a formative or ex-ante evaluation.
- Diversity of programme owners: The agent commissioning an evaluation and responsible for its implementation may become a conflicting issue between the different ministries responsible for programmes within a portfolio.
- Diversity of stakeholders: At the same time as the diversity of programme stakeholders rises evaluation stakeholder variety rises. Stakeholders might be the same but there might also be new stakeholder groups emerging for evaluation or certain groups might have a higher stake in evaluation than in the programme itself.
- Parallel time frames: Not only an ex-post, a formative or an ex-

1 Portfolios are also referred to as funding ecologies (Van den Besselaar/Sandström 2013).

ante-perspective is expected, but often two or even all three perspectives are to be addressed in one evaluation. This is reflected by the time frames the evaluation takes into account.

- Number and diversity of methods and instruments: The more diverse the evaluation objects, the more and diverse the evaluation methods and instruments that have to be tailor-made for each specific evaluation question to be answered.
- Variety of evaluation goals: The more complex the evaluation regarding its function, number of stakeholders, and time frames, the more complex the goals of the evaluation itself become.

CASE STUDY

Portfolio evaluations in a strict sense are still rather rare. A case study of a programme evaluation currently carried out will be used to illustrate the different requirements. This case study is the "**founder contest ICT innovative**"[2] which is being organised by VDI/VDE-IT GmbH on behalf of the German Federal Ministry of Economics and Technology. Even if the case study is more a "quasi" portfolio run by one single programme owner and consisting of a core measure it comprises the main elements of a portfolio. There are parallel but interdependent measures which could also work as standalone instruments as well as plurality of role models and functions of the evaluation itself. The case study therefore is a single programme holding the characteristics of a portfolio because it consists of a portfolio of instruments.

FOUNDER CONTEST ICT INNOVATIVE

The **"founder contest ICT innovative"** comprises a portfolio of different instruments that are directed at different target groups and follow hypotheses that are to some extent interconnected but also to some extent independent of each other. Generally, the founder contest aims at supporting start-upstart-up companies, which focus on products and services in the information and communications technology (ICT) sector. It is organised as a biannual contest that awards considerable prizes as seed money for start-ups (BMWi 2013). Goals of the contest are short- to medium-term to raise the number of start-ups in the ICT industry as well as to help start-ups to be more successful. Long-term goals are a better exploitation of start-up potential in ICT, to create employment in a thriving industry and a contribution to more entrepreneurial spirit in Germany altogether.

The portfolio of instruments of the "**founder contest ICT innovative**" comprises the following:

- All participants of the contest receive an individual written feedback on their business idea.
- The winners are entitled to a set of measures which include the prize itself of up to 30.000 Euros, individual coaching days from professional experts, the chance to participate in an individually conducted strategy workshop, the participation in workshops and seminars covering relevant aspects of founding an ICT-company.
- Another activity is the organisation of a public award ceremony in which the winners are being honoured. Sometimes the award is presented by the German Federal Minister of Economics and Technology in person. This event is widely announced and covered in print media and – more importantly with respect to the target group – in online media.
- Another activity is organising a congress for young ICT companies in Berlin which aims at networking for the community.

Figure 1 shows the expected results of the founder contest on the three different dimensions: output, outcome, and impact. Activities car-

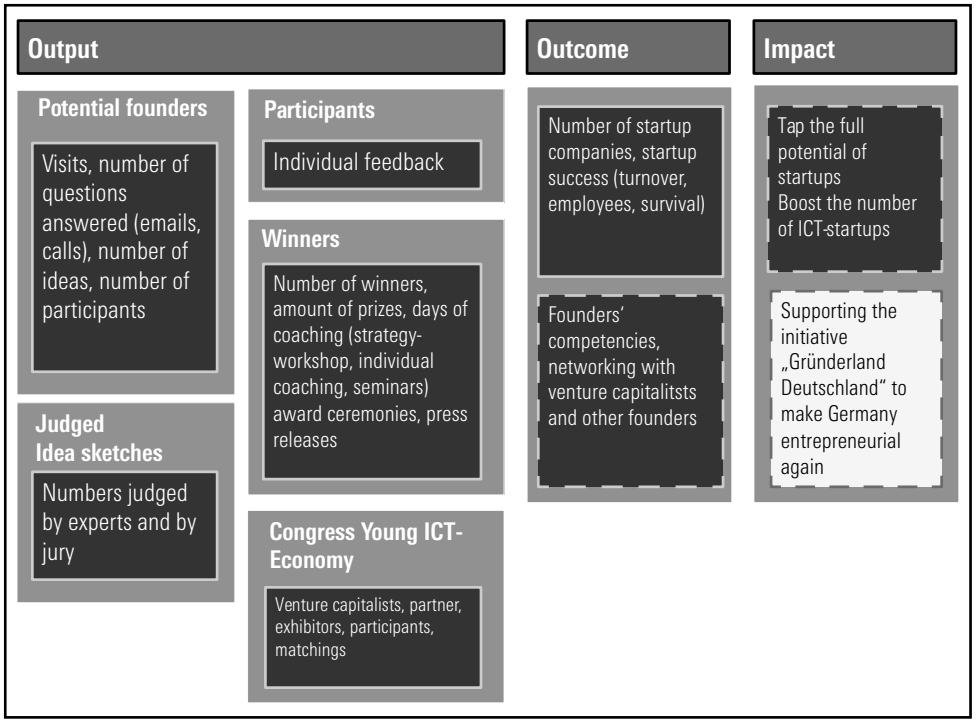


FIGURE 1 - Results of the founder contest

2 "Gründerwettbewerb – IKT Innovativ" (www.gruenderwettbewerb.de)

ried out are the organisation of the contest itself, which means receiving sketches on business ideas, evaluating their quality and finally choosing the winners.

The programme theory behind this portfolio can be divided into five different areas in which specific hypotheses on causal relationships on the impact paths can be identified. The hypotheses cover causal pathways on feedback, competencies, financing, public relations, and networking. Overall is it lead by the central hypothesis that a bundle of measures organised by a single organisation and specifically adjusted to the start-up at hand will have a greater impact than single measures or single measures organised by different agencies.

IMPACT ASSESSMENT AND FORMATIVE EVALUATION OF THE FOUNDER CONTEST

The impact assessment and formative evaluation of the founder contest is being carried out by VDI/VDE-IT's Institute for Innovation and Technology (iit). iit is a separate organisational unit within VDI/VDE-IT. The evaluation team is not involved in organising the contest itself but benefits from close organisational links with the management team, which makes access to data easy and the evaluation efficient.

Reflecting upon the complexity of the programme's approach, the ongoing evaluation is differentiated to reach different goals, to cover different time frames, to address a set of stakeholders and recipients of evaluations results, etc. It covers five main tasks which will be illustrated in more detail:

1. longitudinal study of the participant's activities
2. portraits of successful companies as positive role models
3. concurrent survey to improve the contest
4. scanning developments on the context of the contest
5. mid-term evaluation on effectiveness and efficiency

LONGITUDINAL STUDY OF PARTICIPANTS' ACTIVITIES

The basis for impact assessment is a longitudinal study of the participant's activities as well as an analysis of data obtained from the contest to assess how it affected (and supported) the start-ups' behaviour. For this task a panel has been set up, which allows for an in-depths quantitative analysis of the long-term development of the participant's start-ups (Kerlen/Wangler/Wessels 2013). Once a year, participants are asked to give feedback on the development of the founded companies. With this data, the indirect effects of participation in the founder contest can be shown. The panel is a cornerstone of the summative evaluation at the end of the programme. This element is essential in order to account for the effects of the intervention. Clients are not only the agent commissioning the contest, but also the wider public and auditing authorities within the ministry.

PORTRAITS OF SUCCESSFUL COMPANIES AS POSITIVE ROLE MODELS

The quantitative approach of the evaluation is complemented by a more qualitative access to information. Some of the more successful companies are portrayed as positive role models for other founders. So far the evaluation team realised eight standardised case studies to cover the different background situations at the beginning of the process, the internal as well as the external success factors and especially the continuation of the founding story after participation in the contest. The information was collected by in-depth interviews complemented with data from proposal, surveys, as well as internet and edited in case study format. The field work for this part of communication work is done by the evaluation team, because it is interconnected with the qualitative research into obstacles and success factors of founding businesses in this industry. The evaluator has to be aware of the trap that might arise from the fact that successful founders are being looked at to find success.

CONCURRENT SURVEY TO IMPROVE THE CONTEST

A concurrent survey allows comprehensive, timely and targeted feedback of the contest (Kerlen et al. 2012; Kerlen/Wiedemer/Eckardt 2012; Kerlen/Eckardt 2013). In this survey, the experiences of the participants are collected, summarised and analysed – with the aim of improving the support given to the contestants while further developing the programme's contents and approach. The main instrument to collect data for a yearly adaption of the measure is a survey to all participants. Items covered include feedback to procedures of participation, benefits of participation, main problems in starting a company/ reasons not to start a company, characteristics of start-ups. It provides critical judgments about the funding procedures, but also important insight in the perception of the founders' environment as well as in new trends and developments.

SCANNING DEVELOPMENTS ON THE CONTEXT OF THE CONTEST

Informing all involved actors of the above-mentioned tasks and also producing own insight is the task of scanning developments on the context of the contest. For example, a longitudinal analysis of all German and the major European start-up, business plan, and entrepreneurial development contests is part of this work package. Secondary, analyses and interviews are the methods mostly employed in this task. But there is also one set of questions included in the concurrent survey that focuses on a specific topic, like internationalisation aspects, female entrepreneurs or new models of financing start-ups by crowd funding.

The measure itself had continuously to be argued against similar activities of other players of the innovation system, to prove that there was no redundancy and still a need for an intervention by the German Federal Ministry of Economics and Technology. Due to the very dynamic development of the policy in favour of start-ups on national and regional level, the scanning of the developments had to be realised systematically, even to adapt the measure to new trends and developments.[3] The measure itself was seen as one step in a chain of different support measures, so the matching and connection capability had to be maintained by adaption to environmental changes.

MID-TERM EVALUATION

The mid-term evaluation was realised in a period when continuation of the measure was potentially to be decided due to the end of the legislative period. The results were meant to document causal relationships between the specific instruments and the expected effects like closer networks between start-ups and potential financing institutions or participation of start-up teams and the success of their new-born company (Kerlen/von Drachenfels/Wangler/Wessels 2013). The data used for this exercise was mostly collected by the surveys already introduced.

3 For examples on outputs of this task see Kerlen et al. (2012a); Kerlen et al. (2012b); Kerlen/Wangler/Wessels (2013).

CONCLUSION: CHALLENGES FOR EVALUATION

Main challenges of a portfolio evaluation are to select best empirical access to different measures, to collect relevant data as well as to aggregate individual evaluation results. It is necessary not to eclectically select favourable data to prove success, but to find an approach to define appropriate indicators and collect all relevant data. An annual survey which collects the perceptions and assessments of the target group towards the different measures offered by the portfolio programme leads to comparative data about these measures. The case studies give access to an assessment of the interdependence of these measures, to show the patterns of use by the target group itself of those instruments.

The evaluation also looks at independent influencing factors and additional external measures of the programme "environment" in a mid-term perspective. By this, a greater picture of publicly funded support programmes and instruments for start-ups in Germany can be designed and the core object of the evaluation (the programme itself) can be compared to other measures. From the viewpoint of the target group, the boundaries between the evaluated programme and other measures are rather artificial; the decision about using a supportive opportunity is not taken on the basis of formal ownerships of ministries or funding agencies. The start-ups expect the German government to design a systemic supporting scheme where different measures interact in the best possible way.

Important for a success of this procedure is also the right role for the evaluation team. It should, on one hand, act as an independent evaluation unit with a neutral view of the process to be evaluated. The use of high quality survey data makes sure that the perspective of the target group is included in the evaluation process. The evaluation team should, on the other hand, support the programme owners and the programme agency in further developing the measure. Because of the trustful interaction with the programme agency, the evaluation team can realise a critical but constructive discussion also based on internal process details of the measure implementation. The evaluation team therefore acts in some respects as a coach for the programme agency.

This double role asks for double competencies of the evaluation team members. Furthermore, considering the use of a mix-method approach, a team is needed with qualifications spanning from qualitative research to multivariate analysis, complemented by expert knowledge in ICT industry with experience as an independent opinion maker as well as facilitator and coach. Methodological competencies in evaluation have to be combined with expertise in the area of start-up support to be able to participate in a qualified discussion with the programme management about specific aspects of the programme and fulfil the role expectations as a coach. In this setting with different expectations by stakeholders, the role clarification of the evaluators for themselves and for others becomes essential.

Complicatedness and complexity of portfolios correspond to a raising complexity of portfolio evaluations. To find proper answers to the resulting evaluation demands is not only true for portfolio evaluations. But it is a showcase highly suitable to illustrate the challenges.

REFERENCES

Bundesministerium für Wirtschaft und Technologie [Hrsg.] (2013): Gründerwettbewerb – IKT Innovativ. Informationen zum Wettbewerb. Stand April 2013.

Fischl, Iris/Kulicke, Marianne/Wessels, Jan (2013): Portfolioevaluierungen: Herausforderungen und aktuelle Erfahrungen. In: Zeitschrift für Evaluation 2/2013, 338-341.

Kerlen, Christiane/ von Drachenfels, Christian/ Wangler, Leo/ Wiedemer, Volker (2012a): Bestandsaufname zur Situation von IKT-Gründungen im März 2012 unter besonderer Berücksichtigung des Gründerwettbewerb – IKT Innovativ. März 2012.

Kerlen, Christiane/ von Drachenfels, Christian/ Wangler, Leo/ Wiedemer, Volker/Eckardt, Iris (2012b): Bestandsaufname zur Situation von IKT-Gründungen im September 2012 unter besonderer Berücksichtigung des Gründerwettbewerb – IKT Innovativ. September 2012.

Kerlen, Christiane/ Wangler, Leo/ von Drachenfels, Christian/ Wiedemer, Volker/ Brandt, Claudia (2012): Zur Wirkungsanalyse des Gründerwettbewerb – IKT Innovativ. Analyse der Befragungsergebnisse des Wettbewerbsjahrgangs 2010. März 2012.

Kerlen, Christiane/ Wiedemer, Volker/Eckardt, Iris (2012): Zur Wirkungsanalyse des Grün-derwettbewerb – IKT Innovativ. Analyse der Befragungsergebnisse des Wettbewerbsjahrgangs 2011. September 2012

Kerlen, Christiane/ Eckardt, Iris (2013): Zur Wirkungsanalyse des Gründerwettbewerb – IKT Innovativ. Analyse der Befragungsergebnisse des Wettbewerbsjahrgangs 2012. Oktober 2013.

Kerlen, Christiane/ von Drachenfels, Christian/ Wangler, Leo/ Wessels, Jan (2013): Zwischenevaluation des Gründerwettbewerb – IKT Innovativ. Paper presented at 17th Annual Interdisciplinary Entrepreneurship Conference, 7-8 November 2013 in Koblenz.

Kerlen, Christiane/ Wangler, Leo/ Wessels, Jan (2013): Aktuelle Daten und Trends zur IKT-Gründungsförderung in Deutschland. August 2013.

Van den Besselaar, Peter/Sandström, Ulf (2013): Evaluation at the research systems level: Funding ecologies as policy portfolio. Paper presented at Evaluation of STI policies, instruments and organisations: new horizons and new challenges Conference, November 14, 2013 – November 15, 2013 in Vienna.

AUTHORS

CHRISTIANE KERLEN
Dr Kerlen Evaluation
E: info@kerlen.co.uk

JAN WESSELS
Institute for Innovation and Technology
E: wessels@iit-berlin.de

VOLKER WIEDEMER
Hochschule Magdeburg-Stendal
E: volker.wiedemer@hs-magdeburg.de

THE EUROPEAN RESEARCH AREA: A PROCESS-RELATED CHALLENGE FOR INDICATOR AND POLICY DESIGN

A POLICY REFLECTION ON HOW TO ADVANCE ERA BASED ON RECENT EVALUATION LITERATURE

MATTEO RAZZANELLI[1]

INTRODUCTION

This paper proposes a policy reflection on the approach followed by the European Union (EU) in developing its policy with regards to national R&D policies. As of November 2013, the European Research Area (ERA) initiative is about steering a variety of actors in the European R&D policy space towards the fulfilment of EU objectives (integration, efficiency, inclusiveness) by defining specific goals for each actor.

The paper uses concepts created by the scholarly community, mainly in the fields of science, technology and innovation (STI) evaluation and indicator design, to make observations on current challenges to the realisation of a European research area. The reason behind pointing out such challenges is to make the realisation of the European research area more likely, given that many players at all levels share a firm conviction that Europe needs it.

Firstly, based on literature about the role of indicators, the paper argues that the Open Method of Coordination approach (2003-2008) was discontinued ultimately because of a lack of progress indicators. Secondly, based on evaluation literature, the paper argues that the current indicator-based ERA Communication approach (2012-2014) does not take into account the distributed nature of R&D policy knowledge and governance. As a consequence, the paper argues that advancing ERA in the future will depend on the existence of a policy design, implementation and monitoring process that is able to solve the shortcomings of its predecessors, while retaining their strengths.

In sum, ERA poses a challenge to policy-makers and experts that is, first and foremost, process-related[2].

The paper is organised as follows: the first section provides the basic definitions and scope for the paper, limiting it to R&D policy at EU and national levels. The second section provides an account of ERA, understood as EU policy on national R&D policies. The third section reviews a few contributions from the literature that are considered relevant to analyse current ERA approaches. Based on section three, section four proposes a reading of the ERA developments found in section two. The section concludes that advancing ERA means finding a suitable participatory process for policy design. Following the analysis in section four, the fifth section proposes a possible way forward in analysing recent indicator design processes that could hold lessons for participatory indicator-design processes used as vectors for interaction and policy design.

DEFINITIONS AND SCOPE

Arguments made in this paper refer to R&D[3] policy. R&D policy is defined here following Wintjes and Nauwelaers (2007, p. 8): "policies that intentionally aim to affect the behaviour of R&D performers, changing size, scope, timing and content of R&D activities by public and private R&D performers". In developed countries, such policies are enacted by funding R&D-performing individuals and institutions (both public and private). R&D policy is chosen as a scope for this paper over innovation[4] policy, which is primarily concerned with the activity of diffusing and applying knowledge[5]. R&D policy by contrast is focused on pure knowledge creation and on knowledge creation about applications[6].

Examples of R&D policy are: organisational models for and agenda setting in science funding; procedures for project proposal evaluation (peer review); policies related to R&D human resources; research infrastructures; access to research data and scientific publications. All of

1 Views expressed here are personal and do not represent the view of the Science Europe Member Organisations or of the organisation as a whole.
2 As opposed to being primarily methodological.
3 Research and experimental development as defined by the Frascati manual (OECD, 2002): "creative work undertaken on a systematic basis in order to increase the stock of knowledge, including knowledge of man, culture and society, and the use of this stock of knowledge to devise new applications".
4 Innovation is defined by the Oslo manual (OECD/Statistical Office of the European Communities, 2005) as "the implementation of a new or significantly improved product (good or service), or process, a new marketing method, or a new organizational method in business practices, workplace organisation or external relations".
5 According to the Oslo Manual definition of innovation, the application of existing knowledge in a context where such knowledge was not previously applied qualifies as innovation.

these fit the given definition because they have a direct impact on the behaviour of R&D performers, and directly affect the R&D activities they carry out.

The scope for this paper is limited to the most traditional model of science policy[7] because the ERA initiative was de facto mostly concerned with reviewing and changing national practices specifically with regards to R&D policy[8]. As explained below, the EU has a long history of science funding under the leadership of the European Commission Directorate General (DG) responsible for research policy. However, it only started setting goals for national R&D policies in 2000. EU-level innovation policy has been mainly pursued outside of the ERA initiative, under the leadership of a variety of European Commission DGs such as those dealing with industrial policy, telecommunications and information technologies, regional policy and the single market. Within DG Research and Innovation, as of November 2013, innovation and R&D policy are still dealt with separately, with the exception of the new Framework Programme (FP) Horizon 2020, which integrates both. Since 2011, national public expenditure on research is also addressed within the EU policy process known as European Semester[9].

THE EUROPEAN RESEARCH AREA INITIATIVE

The ERA concept was officially launched by a European Commission Communication of January 2000[10], which was taken up by the Lisbon European Council of 23-24 March 2000 in its Conclusions[11]. Even if the Communication also mentions private investments, the analysis focuses almost exclusively on public investments and policies for research. In the paragraph entitled "Organisation of Research in Europe", the Commission states that the main EU-level policy for research until then had been the European FP. The Communication reminds that FPs represent a small percentage of the total public research investments in the EU, and that the "principal reference framework for research activities is national" (p. 7). The text concludes that "[i]t cannot be said that there is today a European policy on research. National research policies and Union policy overlap without forming a coherent whole. If more progress is to be made a broader approach is needed than the one adopted to date." Essentially, the Commission advocated going beyond pooled public spending on research (the FPs and other intergovernmental initiatives and infrastructures), by creating a EU-level policy devoted to the integration and "decompartmentalisation" of national research policy. In short, a EU policy on national R&D policies.

ERA was then pursued within the broader policy framework of the Lisbon Strategy for growth and jobs launched by the March 2000 European Council. For those policy areas where the Strategy gave the EU a role in the absence of a Treaty competence, the Council introduced the Open Method of Coordination (OMC)[12] as the main methodology for policy design and implementation. In practice the OMC entailed regular meetings of government representatives and Commission officials to discuss specific policies, followed by reports and other documents related to the subject matters discussed. In the case of research policy, the Commission followed up by issuing an "action plan" for investing in research[13]. The action plan aimed at developing "a common understanding at all policy levels and by all stakeholders" and ensuring progress via the OMC. In terms of content, the plan contained elements of both R&D and innovation policy, with actions comparable to the commitments found in the 2010 Innovation Union Flagship Initiative[14]. The action plan kick-started four OMC "cycles" from 2003 to 2008, carried out by CREST, a permanent advisory body on research policy composed of government representatives and Commission officials[15]. These cycles focused on research policy (rather than innovation) and participation was very much limited to ministerial representatives[16]. Towards the end of the Lisbon Strategy, the OMC in research policy was assessed by a European Commission expert group[17]. The group found the OMC to be good in terms of mutual learning, but underachieving in terms of policy coordination[18]. The overall goals were found to be unclear to both participants and to the national civil servants who should have been the users of OMC outputs[19], such as guidelines and best practice documents.

6 Medicine and engineering sciences for example mostly focus on knowledge creation about applications, therefore practicing them can be considered R&D activity. By contrast, the actual diffusion and user adoption of engineering and medical applications is addressed by innovation activities and policies.

7 See Arnold, 2004, p. 10: this model responds to the traditional rationale for public research funding (public compensation for private underinvestment in knowledge creation, which is a public good). On this topic, see also Martin and Tang, 2007.

8 For the current period, see ERA Actions contained in COM(2012) 392 final. For the period between 2003 and 2009, ERA policy was based on CREST activities. A report on these activities states: "CREST-OMC is focused only on research policy topics" (European Commission, 2009a, p. 4).

9 See http://ec.europa.eu/europe2020/making-it-happen

10 COM(2000) 6 final of 18/1/2000 entitled "Towards a European research area".

11 http://www.consilium.europa.eu/uedocs/cms_data/docs/pressdata/en/ec/00100-r1.en0.htm

12 The OMC was described as consisting in: fixing guidelines for the Union combined with specific timetables for achieving the goals which they set in the short, medium and long terms; establishing, where appropriate, quantitative and qualitative indicators and benchmarks against the best in the world and tailored to the needs of different Member States and sectors as a means of comparing best practice; translating these European guidelines into national and regional policies by setting specific targets and adopting measures, taking into account national and regional differences; periodic monitoring, evaluation and peer review organised as mutual learning processes. Source: Council of the European Union, Presidency Conclusions, Lisbon European Council, 23-24 March 2000, 19 June 2009, Nr: 100/1/00.

13 COM(2003) 226 final/2

14 COM(2010) 546 final of 6/10/2010

15 CREST's mission has been changed in 2010 and its successor is called European Research Area Committee (ERAC).

16 European Commission, 2009, p. 4.

17 The assessment was published in January 2009 (see European Commission, 2009a).

18 European Commission, 2009, p. 31.

19 "The lack of clarity about the nature of CREST-OMC renders the goals and final purposes of CREST-OMC opaque to the final users of its results, namely all national civil servants dealing with research (and innovation) policies (not only those involved in the process)" (European Commission, 2009, p. 32).

In December 2009, EU policy on research switched gears. With the entry into force of the Lisbon Treaty, the EU was given an explicit competence not only in supporting the European scientific and technological base through the FPs, but also in steering national research policies by "achieving a European research area in which researchers, scientific knowledge and technology circulate freely, and encouraging it to become more competitive, including in its industry, while promoting all the research activities deemed necessary by virtue of other Chapters of the Treaties."[20]

The revised Treaty therefore widens the mandate for EU institutions in research policy and formally gives them the task to go beyond pooled funding. This new mandate was pursued via the Europe 2020 Strategy[21], which was articulated in seven Flagship Initiatives. One of these is the Innovation Union, consisting of fourty-two commitments, the issue of a communication on ERA being one of them.

The communication on ERA[22], issued in July 2012, focuses on public R&D policy, leaving innovation policy to other Innovation Union commitments[23]. It can therefore be argued that the ERA Communication marks a de facto second attempt by EU institutions at a European-level policy designed to affect, coordinate or change national research policies in view of achieving supra-national policy goals.

The approach followed in this second policy cycle is different from the one followed under the Lisbon Strategy. To address the concerns of Lisbon Strategy critics[24], the Europe 2020 strategy put a strong emphasis on commitments, targets, monitoring and measurability. The strategy has five overall quantitative targets, and its Flagships also contain indicators to monitor progress.

The ERA Communication follows this logic in that it consists of five strategic objectives ("priorities"), to be achieved through a list of sixty actions assigned to three policy levels (European Commission, Member States, Stakeholder Organisations[25]). The Communication then requires continuous monitoring of policy implementation via "a robust ERA monitoring mechanism (EMM) based on indicators for all the actions to monitor ERA policy reforms and their implementation"[26]. The EMM is to be used for the production of annual reports informing policy decisions by EU institutions (ERA Progress Reports).

In terms of implementation, the ERA Communication calls for a "partnership". Member States and the advisory bodies formed by Member State representatives are considered as "primary actors", whereas other stakeholders (such as national research agencies) as implementation agents (pp. 14 and 15 of the Communication).

Compared to the OMC, the ERA Communication therefore does away with mutual learning aspects and opts for top-down policy design. The Communication provides a list of fixed objectives and requires national policy players to implement them.

Finally, the arrival of the EU on national policy spaces has also entailed a series of efforts by European Commission services to gather data and accumulate knowledge on national research systems. This has been done via a series of studies and data collection contracts[27].

LITERATURE CONTRIBUTIONS RELEVANT TO THE ANALYSIS OF CURRENT ERA POLICY DEVELOPMENTS

Some of the concepts developed in the field of STI evaluation and indicator design seem particularly relevant to the analysis of the EU policy development sketched above. In particular, two contributions can be singled out: (1) a model of public research systems; (2) the understanding of indicators as social constructions.

The first aspect refers to the understanding of public research systems as multi-actor and multi-level interaction spaces (Lepori, 2011). Regardless of formal provisions, actors in these spaces (such as researchers, universities, research councils, ministries) are largely autonomous: each actor legitimately pursues its own goals, and in so doing interacts with other actors in the same space. This kind of modelling abandons the idea of outright top-down policy steering by the government. On the contrary, even government decisions are seen as endogenous to the system. Like all system outputs, government decisions are also affected by organisational structures and interaction patterns within the system. Furthermore, a variety of coordination modes between actors co-exist in the public research system. Coordination modes are rules and patterns governing interactions, such as hierarchies, markets (competition), or informal human networks. Specific coordination modes are used by the government to make sure that independent actors produce those public goods that justify the public funding they receive.

The multi-level nature of public research system refers to the fact that actors in the system perform different functions, which range from providing the overall funding to performing research activities. Depend-

20 Consolidated Version of the Treaty on the Functioning of the European Union art. 179, 2008 O.J. C 115/47.
21 COM(2010) 2020 final of 3/3/2010.
22 COM(2012) 392 final of 17/7/2012. The definition of ERA given therein (p. 3) is: "a unified research area open to the world based on the Internal Market, in which researchers, scientific knowledge and technology circulate freely and through which the Union and its Member States strengthen their scientific and technological bases, their competitiveness and their capacity to collectively address grand challenges."
23 Based on the above definition of R&D policy, twenty-five commitments (59%) can be classified as related to innovation rather than science policy: 7, 10, 10 (RSFF), 11, 12, 13, 14, 15, 16, 17, 17 (cont), 18, 19-A, 19-B, 21, 22, 23, 24-25, 26, 27-A, 27-B, 28, 33, 34-A, 34-B. ERA does include some policy actions that can be considered as part of the grey area between R&D and innovation policy, namely those related to knowledge transfer.
24 See http://ec.europa.eu/europe2020/services/faqs/index_en.htm
25 In practice, organisations with a public mission, such as research funders and performers, universities included.
26 COM(2012) 392 final, p. 14 and also Annex 7 of the ERA impact assessment, European Commission Staff Working Document SWD(2012) 212 final of 17/7/2012.
27 For a list, see: http://ec.europa.eu/research/innovation-union/index_en.cfm?pg=other-studies

ing on how systems are organised nationally, individual organisations can perform different functions and therefore be placed on one or more levels. Figure 1 provides the representation of public research systems proposed by Lepori, together with the functions attributed to each level.

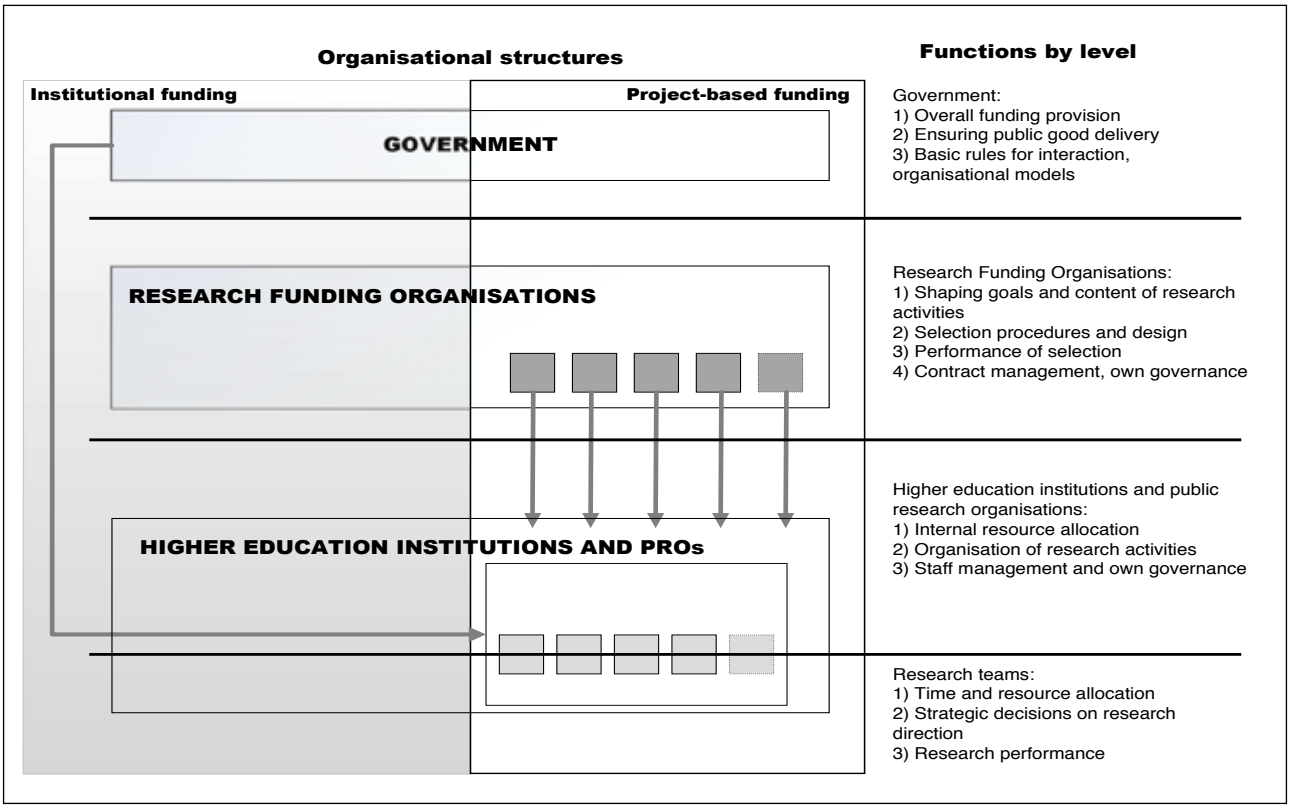


FIGURE 1 - A model of public research systems. Source: diagram on the left-hand side is a reproduction, with slight adaptations, of the diagram published in Lepori, 2011. The column on the right-hand side, "Functions by level", is added to the diagram, based on the texts in Lepori, 2011 and in Reale et al., 2012.

The second contribution from the literature refers to the understanding of indicators as social constructions. Barré (2004) argues that knowledge is always context-specific and path dependent. Godin (2004, p. 3) holds that statistics are based on considerations which have nothing to do with mathematical science.

Barré concludes that indicators should not be regarded as objective representations of the truth. They have no unequivocal interpretation and remain debatable. They do not necessarily help establish causality. Nevertheless, indicators are used in policy making, and they are useful. Godin (2004) identifies three uses for them: theoretical (to understand S&T-related phenomena), practical (to inform decision-making) and symbolic/political (to convince people of an argument).

One may wonder whether the usefulness of indicators in light of their social nature declines. Barré's answer is no. According to Barré, the social nature of indicators can be used to turn them into a policy tool for an "Agora Model" of policy-making. In this view, indicators can be used as a common language (numbers) to structure and foster dialogue in the research policy space. The criticism and debate around indicators becomes a way to address the questions at stake, and to foster an exchange and a synthesis between competing views. To this, we may add Gault's observation (2011) that available indicators in turn affect the determination of policy objectives, thus making indicators extremely relevant for policy design.

However, Barré (2004) suggests that – to play this role – indicators need to fulfil some specific requirements. Two of these are the Agora Model criterion and relevance[28]. The Agora Model criterion is best re-labelled as social robustness criterion. Social robustness means indicators need to be debated by both stakeholders and experts, and that both need to be part of the production process. To this end, indicator designers need to be transparent on both technical aspects and the assumptions and conceptual links embedded in the choices made.[29] Relevance means that indicators need to embed a deep understanding of the stakes as well as stakeholder and decision-maker needs in the different contexts where the indicators may be used.

28 The third one is reliability, which relates to the methodological robustness of indicators (Barré 2004, p. 129).
29 Barré 2004, p. 129

ADVANCING ERA BY DRAWING ON LITERATURE INSIGHTS ON THE TWO ERA APPROACHES

The above two literature insights can be used to guide a policy reflection with regards to ERA, and in particular with regards to the OMC and then the ERA Communication approaches.

In relation to the literature reviewed above, the OMC shows the following strengths:

1. The learning objective and process recognised the actor-based nature of the research policy space. The fact that mutual learning and discussion were placed at the core of the method means that there was an implicit recognition of the actor-centric nature of research policy, and that the definition of objectives and progress monitoring needs to go through a social, participatory process.

2. The multi-level nature of the process was implicitly reflected by the fact that EU and national goals were placed on equal footing, with reasonable emphasis on the concept of 'policy mix'.

However, the OMC can also be regarded as showing some weaknesses:

3. The final assessment of the exercise conveys a widespread lack of sense of clear purpose and achievement, not just in terms of policy coordination, but also in terms of any concrete outcomes for individual players. This could be attributed to the lack of indicators with a symbolic use.

4. One of the main results of the OMC regarded mutual learning, however it is hard to capture and express the impact of mutual learning when evaluating process outcomes.

5. Even though the actor-based nature of research systems was implicitly acknowledged, the only actors involved were ministerial representatives. This fails to capture the majority of actors in the research policy space, especially in light of the fact that EU ministries delegate a large range of functions to dedicated public organisations, as shown in Figure 1.

In sum, OMC cycles focused on mutual learning and peer review of national practices. Very little was done in the way of using shared indicators to structure policy debates. Discussion for example on national policy mixes could not significantly rely on the theoretical and practical role of indicators as suggested by Godin. In the end, what really seemed to cause the end of the OMC was the lack of indicators with a symbolic use, a lack which made it hard for policy makers to buy into and sell the usefulness of the exercise. It also proved hard for process participants to move from mutual learning to coordination. This could explain the very high emphasis put in the Europe 2020 narrative on progress indicators and monitoring.

On the other hand, the ERA Communication seems to show the opposite strengths:

1. Monitoring, in particular via a list of indicators, is used with a strong symbolic use.

2. The list of sixty actions is intended to bring about policy coordination, and change for those actors whose policies are not aligned to actions.

3. The multi-actor nature of the policy space is captured more broadly, since an "ERA Partnership" is created between the European Commission and national organisations with a public mission present in the research space, thus going beyond national ministries in pursuing ERA objectives.

The ERA Communication can be regarded in light of the literature as showing two main weaknesses:

4. There is no understanding of indicators as vectors for interaction, on the lines of the Agora Model. ERA Partners were only marginally consulted to set actions and indicators, and the process of setting objectives and defining indicators is not constructed as participatory and iterative.

5. The independence enjoyed by stakeholders as emerging from the literature is not reflected in the ERA approach. The ERA Communication clearly defines research policy as a domain where governments are "primary actors", with policy design responsibility, whereas stakeholders are to be used as implementers and data collectors[30]. Indicators in the EMM are therefore used as policy implementation tool, rather than policy discussion and design vehicles.

These two shortcomings can be problematic in light of the fact that the research policy delegation from governments to specialised agencies means that policy knowledge is distributed across actors. If ERA policy design is centralised, a likely consequence is the failure to fully embed the knowledge held within specialised organisations in ERA Actions and EMM indicators. This clearly seems to be the case with some indicators[31]. ERA Actions topics where relevant knowledge and the related governance are distributed across public stakeholders are issues like access to scientific publications or modes for cross-border collaboration.

Secondly, actions and indicators designed by a limited set of policy players cannot be considered as socially robust in the sense proposed by Barré. A consequence in ERA is that the Communication contains no narrative on the role played nationally by stakeholder organisations: ob-

30 COM(2012) 392 final, pp. 14-15.

31 For example, the EMM misinterprets some inter-agency model agreements as intended to foster cross-border collaboration, whereas their actual aim is to cope with emergency cases. The corresponding indicator measures the budget allocated to these agreements, whereas no budget is allocated up front to such agreements.

jectives and indicators are based on the narratives of integration, competition and inclusiveness. The Agora Model seems particularly relevant for issues of high complexity, where needs can only be discovered through collective learning processes. In conclusion, current ERA policy developments examined through the lens of literature contributions suggest the importance of finding a working policy design, evaluation and implementation process.

Whereas evaluation is evolving into a well-developed craft, the challenge rests in achieving social robustness and collective learning about systemic bottlenecks and ways to solve them.

The EU experimented with a participatory approach to policy design (the OMC), to then opt for a top-down policy and indicator design process (ERA Communication), coupled with indicator-based policy implementation (EMM), which does not reflect the multi-level and multi-actor nature of the research policy space, and that as a consequence does not embed vital policy knowledge in goals and monitoring means.

BUILDING A PARTICIPATORY PROCESS FOR ERA POLICY DESIGN[32]

If the above analysis is accepted, then the question is how to structure a working process for policy design using indicators to solve some of the shortcomings of the OMC, and to enhance the strengths of the ERA Communication approach.

A possible source of lessons to be learnt could be recent data-gathering efforts conducted at European level in order to create datasets capturing a wide range of features of European research systems. A rough and preliminary analysis suggests a potential for significant lessons to be learnt. Examples of projects that could be used are:

- JOREP[33], on openness and coordination levels of national programmes as measured in terms of budget allocation: this project provided data that was important for policy discussion around the first ERA Progress Report[34] and can likely play a major role in future debates on research agenda coordination on the input side.
- EUMIDA[35], a census and profiling of European universities: this project helped establish a policy definition of universities as independent actors, rather than simply as higher education providers (Lepori and Bonaccorsi, 2013).
- She Figures[36], on gender balance in science: this Commission publication helps structure the debate around gender balance in science, with a process that includes gathering original data via national contact points.
- MERIL[37], a census of research infrastructures: this dataset is an example of a mostly participatory process where original governance solutions to dataset construction seem to be working[38].
- ESF Indicators of Internationalisation[39], a dataset proposed and built with a fully bottom-up process carried out by a group of national research funding and research performing organisations.

For these five examples, the table below suggests possible aspects to be analysed by future research. For each example, the participatory aspects of the process are highlighted, and a suggestion is made on the lessons that seem to be emerging from exploratory and anecdotal evidence[40].

Possible lessons that could emerge from recent indicator-design processes, and that could be used as a basis for policy learning and design can be summarised as follows[41]:

- The experimental design phase of indicators seems crucial for collective learning and for substantial policy discussions. However, in this phase, it is problematic to work with national statistical authorities (EUMIDA, JOREP).
- For dataset building purposes, centralising concept definition, while decentralising concept application seems a promising model (MERIL, EUMIDA, JOREP).
- Social robustness can impact data quality or coverage (EUMIDA, MERIL), but incomplete geographical coverage (JOREP) or other shortcomings (EUMIDA) can be offset by the value of the new information.
- The sustainability of fully bottom-up processes has not been proven, as different actors have their own strategies and views (ESF Indicators of Internationalisation), but even light top-down steering can go a long way in ensuring sustainability (MERIL).
- Bottom-up processes facilitate the emergence of different narratives compared to the narratives emerging from international strategies (ESF Indicators of Internationalisation seems to respond to a tailor-made narrative compared to other international datasets).

32 The author gratefully acknowledges useful information on STI indicators and JOREP received from Matthieu Delescluse, Policy Officer, European Commission, who contributed in his personal capacity under his own name. The author is also grateful to Peter van den Besselaar and Maud Evrard for sharing their knowledge about ESF Indicators of Internationalisation and MERIL respectively.

33 "JOREP, Joint and Open Research Programmes", contract no. RTD/DirC/C3/2010/SI2.561034.

34 http://ec.europa.eu/research/era/progressreport2013_en.htm

35 "EUMIDA, European University Data Collection", contract no. RTD/C/C4/2009/0233402.

36 http://ec.europa.eu/research/science-society/index.cfm?fuseaction=public.topic&id=1282

37 "MERIL - Mapping of the European Research Infrastructure Landscape": http://www.esf.org/serving-science/ec-contracts-coordination/meril-mapping-of-the-european-research-infrastructure-landscape.html

38 http://www.esf.org/media-centre/ext-single-news/article/new-european-database-launched-to-map-strengths-and-gaps-in-research-infrastructure-provision-and-en.html?utm_source=feedburner&utm_medium=feed&utm_campaign=Feed%3A+esf%2FGlobal_News+%28ESF+-++Global+News%29

39 European Science Foundation, 2012.

40 The last column lists the sources used to draw the tentative lessons suggested here.

41 For the sources used, see Table 1.

Project	Narrative	Indicator use	Indicators on	Concept definition	Process	Participatory aspects	Lesson	Reference
JOREP	Integration	Theoretical, practical	Openness and coordination levels of national programmes as measured in terms of budget allocation (coordination of research agendas, input side)	Driven by experts from universities and public research organisations, and EC	Experts, PROs	PROs participated and provided the data, the project built a pilot dataset. Concept definition was centralised, whereas concept application was decentralised.	1) Difficult to involve NSAs at experimental design phase (staff availability, length of statistical cycles, limited expertise with micro-data requiring subject-matter expertise); 2) The pilot was successful, which shows that (theory-based) indicators for science policy can be built with a participatory process; 3) The pilot dataset was used for the ERA Progress Report and the Innovation Union Report, which shows that dataset limitations such as geographical coverage can be offset by the value of new information	Reale et al., 2013 and discussions with people involved in the project
EUMIDA	Competition	Theoretical	Census of European universities, including their profiling	Driven by experts and EC	Experts, NSAs	Centralised concept definition, decentralised concept application	1) Political controversy impacts on dataset coverage, depth and quality. Census linked to a policy shift: universities as independent and competing actors. NSAs not comfortable with new concepts that are not perceived as 'objective'. This lowered the quality and completeness of the dataset. 2) Centralised concept-definition and decentralised concept application can be a good compromise; 3) Absence of a participatory process impacts on data quality and availability	Lepori and Bonaccorsi, 2013
She Figures	Inclusiveness	Practical, symbolic	Gender balance in science	EC, experts	Policy makers (EC), NSAs, Eurostat, national government officials (national statistical correspondents)	Part of the data is traditional statistics, but part is gathered nationally and aggregated centrally	Consolidated concepts and shared narratives allow for complex processes mixing official statistics and stakeholder data	She Figures publications
MERIL	Integration	Practical	Census of research infrastructures of international relevance	Stakeholders, EC	European stakeholder organisation (ESF), national policy-level players	Participatory process based on central concept definition and decentralised concept application. Involvement of data providers ongoing	MERIL follows previously failed attempts. Previous attempts failed to deliver coverage and data quality, problems with concept definition and application. MERIL solved the problem by designing criteria for inclusion in a participatory way (EC, stakeholders) and then by assigning 'gatekeepers' for criteria application. Results are promising, as about half of the expected entries are in the database. The main challenge to be solved is the awareness and involvement of final data providers, as users seem to be mostly policy makers	Publicly available materials and discussions with people informed on the project
ESF Indicators of Internatio-nalisation	Efficiency in pursuit of national missions	Practical	Performance of Funding Agencies in terms of internationalisation of activities	Experts, stakeholders	Experts, stakeholders	Process fully participatory, from concept definition to data collection	Exclusion of central policy makers means that commitment to dataset depends on individual strategies, interests and resource availability of each organisation. The dataset was conceived as of practical use (to inform strategies), therefore a strong narrative was missing. This questions the sustainability and continuation of the exercise	Reale et al., 2012. Discussions with project players

TABLE 1 - Examples of ERA-related indicator design processes that could hold lessons for future policy design and monitoring approaches, as well as relevant aspects to investigate. EC stands for European Commission; NSA stands for National Statistical Authority.

CONCLUSION

By using STI evaluation literature to build a narrative on ERA approaches adopted over the years, I argued here that the recent history of EU policy shows the usefulness of indicators for both policy design and implementation. Such indicators however need to be relevant and socially robust.

As pointed out by Lepori, Reale and Tijssen, (2011, p. 4) this implies a "conceptual shift from a 'linear' process where indicators proceed from design towards (standardized) production and interpretation towards a more interactive process, where indicators are contextualized and interpreted in an iterative and open way, with the operational risk that there is never a 'final' set of indicators which can produced regularly". In the case of ERA, the point of indicators is to advance policy, and therefore the risk is not relevant. The reason is the one pointed by Barré that quantitative indicators are changing their *raison d'être* in policy design (Barré 2004, p. 127).

Based on the above considerations, the paper argued that the advancement of ERA requires addressing the challenge of coming up with a policy process where policy design is based on collective learning and dialogue between a wide variety of stakeholders, structured around indicator design, in order to ensure a common language and a sense of progress. This could be achieved by looking at examples of indicator design with or without participatory aspects, to draw lessons on how to build working participatory processes. Such processes could then be part of a future ERA approach, thus making sure that ERA advances.

BIBLIOGRAPHY

Arnold, E. (2004), Evaluating Research and Innovation Policy: a Systems World Needs Systems Evaluations, Research Evaluation 13 (1): 3-17

Barré, R. (2004), S&T Indicators for Policy Making in a Changing Science– Society Relationship, in Moed, Glänzel and Schmoch (eds.), Handbook of Quantitative Science and Technology Research, pp 115-132.

European Commission (2009a), The Open Method of Coordination in Research Policy: Assessment and Recommendations - A report from the Expert Group for the follow-up of the research aspects of the revised Lisbon strategy, Directorate-General for Research, Directorate C - European Research Area: Knowledge Based Economy Unit C.3 - Economic analysis and monitoring of national research policies and the Lisbon strategy

European Commission (2009b), ERA indicators and monitoring, Expert Group Report, Directorate-General for Research, Luxembourg: Publications Office of the European Union

European Science Foundation (2012), Indicators of Internationalisation for Research Institutions: a new approach - A report by the ESF Member Organisation Forum on Evaluation: Indicators of Internationalisation, ESF, Strasbourg: FR

Gault, F. (2011), Social Impacts of the Development of Science, Technology and Innovation Indicators, UNU-MERIT Working Paper Series, #2011-008, United Nations University - Maastricht Economic and social Research and training centre on Innovation and Technology: Maastricht, NL

Godin, B. (2004), The Who, What, Why and How of S&T Measurement, Project on the History and Sociology of S&T Statistics Working Paper No. 26, Canadian Science and Innovation Indicators Consortium (CSIIC): Montreal, CA

Lepori, B. (2011), Coordination Modes in Public Funding Systems, Research Policy 40 (3): 355–367

Lepori, B. and Bonaccorsi, A. (2013), The Socio-Political Construction of a European Census of Higher Education Institutions: Design, Methodological and Comparability Issues, Minerva 2013 (51): 271-293

Lepori, B. and Reale, E. (2012), Handbook on Data Collection on Open and Joint Programmes, JOREP European Commission contract number RTD/DirC/C3/2010/SI2.561034

Lepori, B., Reale, R. and Robert Tijssen (2011), Designing indicators for policy decisions: challenges, tensions and good practices: introduction to a special issue, Research Evaluation 20 (1): 3-5.

Martin, B. R. and Tang, P. (2007), The Benefits from Publicly Funded Research, SPRU Electronic Working Paper Series Paper No. 161- University of Sussex: Falmer, UK.

OECD (2002), Frascati Manual 2002: Proposed Standard Practice for Surveys on Research and Experimental Development, The Measurement of Scientific and Technological Activities, OECD Publishing

OECD/Statistical Office of the European Communities (2005),Oslo Manual: Guidelines for Collecting and Interpreting Innovation Data, 3rd Edition, The Measurement of Scientific and Technological Activities, OECD Publishing

Reale, E., Inzelt, A., Lepori, B. and Peter van den Besselaar (2012), The social construction of indicators for evaluation: Internationalization of Funding Agencies, Research Evaluation 21, pp. 245-256.

Reale, E., Lepori, B., Nedeva, M., Duncan, T., Primeri, E., Chassagneux, E. and Philippe Laredo (2013), Investments in JOint and Open REsearch Programmes and analysis of their economic impact (JOREP) - Final Report, European Commission, Directarate General for Research and Innovation, Luxembourg: Publications Office of the European Union

Wintjes, R. and Nauwelaers, C. (2007), Evaluation of the Innovation Impact from Programmes with the Interacting Goals of Research Excellence and Regional Development: How to Decentralize the Lisbon Strategy into Coherent Innovation Policy Mixes? Paper submitted to the International Seminar on Evaluation of Science, Technology and Innovation Policy Instruments, Rio de Janeiro 3-5/12/2007, source: http://www.cgee.org.br/arquivos/SI_ExcellenceandCohesion _WintjesR.pdf.

AUTHOR

MATTEO RAZZANELLI
Science Europe
E: matteo.razzanelli@scienceeurope.org

POSITIONING COMPETENCE CENTRES VIA MONITORING DATA: TOWARDS A SYSTEMATIC APPROACH

MICHAEL DINGES, JAKOB EDLER, MATTHIAS WEBER

INTRODUCTION

Since the mid-1990s Competence Centre programmes have become one of the major support mechanisms in a number of European Union Member States, Australia, Canada and the US in order to ultimately foster industrial innovation through joint knowledge production, training and transfer of researchers and commercialisation of IP[1]. Competence Centre Programmes are funding long-term oriented research alliances between public research performing organisations and industry, performing both fairly fundamental but also more applied problem-oriented research (Arnold et al. 2004). They seek to tackle one of the core challenges of innovation systems: bridging the gap between knowledge production and research outputs of the public research system and the production of market driven and societally valuable solutions in the business community (and broader society). The programmes do not support translation mechanisms between public research and the business community, but the co-production of research agendas and research activities. As such, they are characterised by strong market-oriented strategic research agendas and a close engagement between public research and industry, which does not rest upon the conduct of extensive contracted research, but a focus on truly collaborative research. Over time, competence centres are expected to build core competences in the area of the technology focus of their industrial partners and thereby develop strong linkages between researchers and industry (CREST 2008). Due to their long term nature and comparatively high rates of public subsidy, competence centres are expected to have distinct impacts on joint knowledge creation and circulation, but also on human resources capacity building and internationalisation of research.

Because competence centres have this double mission of creating new organisational structures of cooperation as well as performing concrete research, innovation and training tasks, they necessitate a comprehensive assessment framework, a framework that is able to capture the structural characteristics as well as the performance. This paper provides and applies a conceptualisation to characterise and assess the nature, functions and performance of competence centres. The concept is guided by the idea that the monitoring data that is collected for programme management purposes can support both purposes. Thus, we develop measures and means that draw on programme monitoring data allowing to link centre assessment to the programme level. In doing so, we develop a broader understanding of the purposes and requirement for an adequate monitoring system to guide programme implementation and assessment.

Our approach is motivated by the observation that monitoring data is widely used in European innovation policy evaluation (Edler et al 2012), but that this use is largely descriptive and entirely inadequate to derive at deeper insights into the structures and specificities of the funded entities. In this article we want to show how monitoring data can be used to characterise and typify the structures that are funded, to better understand their function in the innovation system and to allow centre comparisons. By doing so, we add a new dimension of monitoring that incorporates the positioning rationale into evaluation practice. The positioning indicator rationale stresses that for national innovation systems (and programmes) a) the position of organisational actors (their identity, relationships, complementarities and immaterial assets) are as important as formal inputs and outputs, and b) that indicators are needed to support the strategic decisions of organisational actors, which may help improving coordination and system performance (see Lepori et al. 2008).

The concrete example with which we develop and apply our approach is the Austrian Competence Centre Programmes Kplus, building upon data and insights gained in the ex post evaluation of the programme performed in 2012/2013 (Schibany et al. 2013).

In the following, we first provide a brief presentation of key aspects of the Austrian Competence Centre Programme Kplus and highlight the evaluation and monitoring practices of the programme. We then develop and apply the mix of indicators to characterise the programme as well as the centres and actors therein. To do so we use programme performance data of the 17 Austrian Kplus centres, and show how the structural data of the Competence Centres allow building a differentiated typology of centres and a positioning of individual centres in the system. This is followed by a discussion of the appropriateness of the indicators used in the existing monitoring system to assess S&T performers (centres) for those who need to assess the performance of the institutions such as governing boards, policy decision-makers (see Lepori et al. 2008). We conclude

1 For a survey on competence centre programmes in Europe see: the 2005-2010 ERA-NET project COMPERA has linked 17 competence centre partner programmes from 12 European countries, exhibiting the significance of competence centre programmes in Europe (http://ec.europa.eu/research/fp7/pdf/era-net/fact_sheets/fp6/compera_en.pdf#view=fit&pagemode=none). For earlier competence centre programmes see Arnold et al. 2004. The 2005-2010 ERA-NET project COMPERA has linked 17 c ompetence centre partner programmes from 12 European countries, exhibiting the significance of competence centre programmes in Europe (see http://ec.europa.eu/research/fp7/pdf/era-net/fact_sheets/fp6/compera_en.pdf#view=fit&pagemode=none).

with more comprehensive suggestions with regard to the development of a monitoring system that serves the multiple purposes of programme implementation and assessment.

THE AUSTRIAN COMPETENCE CENTRE PROGRAMME KPLUS

The Austrian Competence Centre Programmes "Kplus" has been launched in the late 1990s with the objective to raise science industry linkages in Austria to a new level and contribute to the internationalisation of R&D. In 2007 the programme was transferred into a new competence centre programme COMET, together with the slightly more application oriented competence centre programme K_Ind/K_net. This paper focuses on the Kplus programme.

The target groups of the Kplus programme were industrial enterprises and universities and non-university research institutions carrying out high-quality research in fields with high potential for application. For operating the programme, 17 physical Kplus centres were set up, with locations spread almost all over Austria. Funding of the initial Kplus programme provided an annual budget in the range of 2-4 million Euros per year. For each centre, a maximum of 60% was provided by public sources (national and regional funding). Industry was requested to carry the bulk of the remaining funding, both with financial contributions and in-kind contributions, whereby the latter should not exceed 50% of total industrial contribution. Up to 5% was provided by the participating research organisations and universities.

The competence centres within Kplus were established as formal networks with a legal framework (Ltd. Company), which provided easy access for new companies to join in (also for a limited period of time). In terms of geographic scope, all centres aimed to act as national knowledge hubs for companies throughout all federal provinces of Austria, but they were nevertheless strongly embedded in the regional innovation systems. On average about 40% of companies co-operating with the centres were local, innovative SMEs.

As the centres were also seen as a tool for internationalisation of R&D, they were encouraged to co-operate with international companies. To a large extent, the ex-post evaluation of the Kplus programme showed that these cooperating firms stem from the neighbouring country Germany (80%) and Switzerland (8%) building upon existing co-operations which intensified during the operation of the programme.

The centres were further requested to perform not only research within the framework of the Competence Centre Programme – which provided core funding – but also to engage in a) national thematic R&D programmes, b) international collaborative R&D projects, and contract research for companies.

The Competence Centre Programme was one of the earliest Austrian R&I support programmes with continuous monitoring and evaluation. In particular Kplus has been praised for having pioneered the use of evaluations (see OECD 2004, Edler et al. 2004, Biegelbauer 2008). Already back in 1999, a comprehensive assessment scheme for the programme and the individual centres was set up (Ohler and Stampfer 1999, TiG 2000). At the level of centres, the assessment cycle contained a) a two-stage selection process based upon clear funding criteria, b) a permanent monitoring system, c) an interim assessment, making use of international peers and d) a final assessment. At the programme level a continuous monitoring of additionality effects (Steyer 2006, Steyer et al. 2007, Schmidmayer et al. 2010) and a mid-term programme assessment (Edler et al. 2004) was performed. In order to fully close the evaluation cycle of the Competence Centre Programmes, an ex post evaluation of the programmes was performed in 2012/2013 (Schibany et al. 2013).

MEASURES AND MEANS TO CHARACTERIZE COMPETENCE CENTRES: TOWARDS A TYPOLOGY OF CENTRES

Competence centre programmes show a considerable degree of variation in terms of rules for implementation, operationalization, and (prescribed) structures including a) the location of competence centres (e.g. physical centres vs. virtual networks), b) the characterization of networks with prerequisites to comprise certain partners, c) the funding structures, the governance structures, and d) the selection procedures of centres (cf. CREST 2008).

While the above mentioned variations tend to be the results of rules and requirements imposed by programme owners, we can build a typology of competence centres based upon the resulting structural characteristics and the roles and functions of competence centres within the innovation system. This is at the core of the basic idea of positioning centres. Hence, from a more functional perspective, the following dimensions can be considered:

- **Strategic orientation:** Science driven centres vs. industry driven centres. Centres can be differentiated by their main objectives and composition of partners. Centres can rather be geared at creation of scientific „breakthroughs"and long-term strategic oriented research, or rather focus on the provision of technological solutions for companies.
- **Degree of heterogeneity of actors**: Centres can be differentiated by their inter-sectoral composition of industrial partners and the composition of scientific partners. Furthermore, the regional outreach concerning partners of centres can be taken into account.
- **Degree of internationalisation:** Centres can be differentiated by the degree of internationalisation of its industrial partners and scientific partners.
- **Governance and self-conception**: Centres can be differentiated by their internal and external governance structures (e.g. composition of boards, decision making processes, liability structures). Centres are either platforms used by (industrial) partners on a short term basis, or they can be characterised as own research entities, with clear strategies, visibility and autonomy as regards strategic decisions.
- **Characterisation of centres in terms of size and novelty of cooperation** (i.e. the creation of new networks vs. amendment

of an existing network and its ambition to address new fields and create new combinations of actors).

To characterise centres along those dimensions, two types of data could be mobilised. First, data from the monitoring system of the programme itself - where relevant indicators for our characterisation were integral elements of the monitoring system - and data from databases that were linked to the monitoring data of the Competence Centre programme. Additional qualitative variables, which were not included in the structural data of the competence centres, were retrieved from the mid-term evaluations of the centres, the self-assessment reports (core documents) of the centres and interviews. These data referred in particular to:

- the strategic orientation of the centres in terms of degree of innovation (novelty of research questions tackled by centres and respective results) and time horizon of research streams pursued by centres;
- governance and self-conception of centres, ranging from platforms with limited numbers of employees to strong institutions with several key positions for shaping agendas and activities. Variations observed referred to overall established governance mechanisms of centres such as the involvement of stakeholders concerning strategic research development and scope of activities pursued by centres and the internal governance structures comprising plans for human resource development and research strategy development;
- the characterisation of networks, detailing the novelty of cooperation and the development of networks over time.

Table 1 below shows the indicators for our main dimensions to characterise the centres.

COMPETENCE CENTRE DIMENSION	INDICATORS
Strategic-Orientation	• Share of scientific partners (institutions) engaged in the competence centre programme • Nr. of peer-reviewed publications per million Euro budget • Nr. of patents per million Euro budget • Nr. of PhD and Master theses per million Euro Budget

Degree of heterogeneity of actors	• Scientific heterogeneity
	  • Representation of different scientific disciplines within a competence centre measured by Simpson's diversity index
	• Industrial heterogeneity
	  • Representation of different industrial branches measured by Simpsons's diversity index
	  • Share of large companies
	  • Share of high-tech and medium-high tech manufacturing companies
	• Regional focus
	  • Share of partners stemming from the core region
	  • Degree of representation of Austrian regions

Degree of internationalisation	• No. and share of international business partners • No. and share of international science partners • Participation in EU-FP programmes (no data at centre level available)
Governance and self-conception[2]	• No. and share of employees • No. and share of key researchers
Characterisation of networks	• Annual budget/no. of employees • Annual budget/no. of partners • Novelty of cooperations

TABLE 1 - Towards a typology of competence centres: dimensions and available quantitative indicators for characterisation

2 Additional qualitative indicators retrieved by interviews and interim-evaluation documents of centres related to the positions of management staff, internal responsibilities and authority structures, and relations to partner and host institutions.

The available monitoring data resulted in distinct profiles of competence centres, which allow inter-centre comparison and provide a useful evidence base for positioning the competence centre and reflecting their role in the national innovation system. As an example, two competence centre profiles comprising the quantitative indicators are portrayed in the figure below.

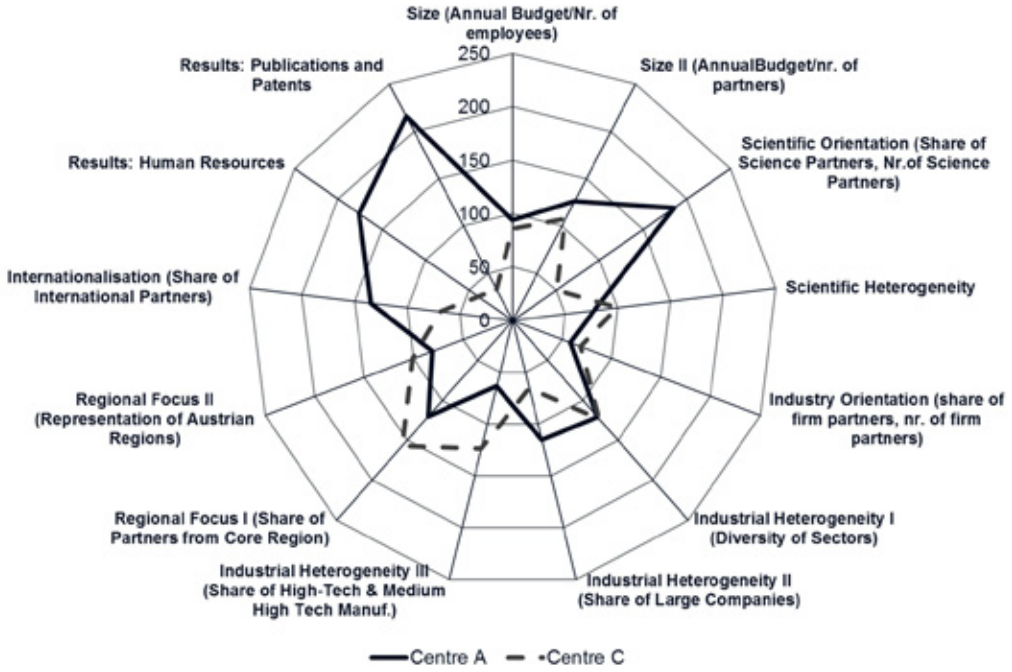


FIGURE 1 - Example profile of Kplus Competence Centres (deviations from average centre (=100)
Source: own illustration based upon Kplus monitoring data

The characterisation of centres further allows both a *differentiated* picture of the landscape and the definition of structural patterns and types of centres. We first see that a programme with clear, homogenous objectives delivered competence centres exhibiting huge diversity in terms of partner structures, specialisations, business models and performance patterns. Second, however, the characterisation allows synthesising the heterogeneity of the centres into a small number of centre types:

- **Regional platforms**, i.e. centres with strong roots in the regional innovation system in terms of partners and focus on needs of technology-oriented SMEs located in similar industrial branches. They are often more applicationÐoriented, build upon existing cooperations and seek strengthening those relationships. Since they are generally focused on the needs of the regional economy, they are organised more as platforms than as strong and independent players.
- **New network centres** were able to cluster competences in several federal states, and thereby often break new ground with cooperation between science and industry. Their activities are primarily oriented towards the usefulness for industrial partners, but the activities can also comprise development of scientific competences and building of a corresponding profil
- **Lighthouse type centres** were generally built on established cooperations between science and industry, and, thus, reflected known strengths of the Austrian economy and/or research landscape. A clearly recognisable scientific profile and the development of a new quality of cooperation are the main foci in this case.
- **Centres of the future** aimed for the development and clustering of competences in relatively new and even less established fields in order to strive to become corporative players. Here, various players are working together in the same direction in order to make use of the common synergy potential and to achieve (inter)national visibility.

POTENTIAL AND LIMITS OF MONITORING SYSTEMS

The example of the Kplus evaluation demonstrated that programme monitoring data allowed the construction of profiles of competence centres, which can serve as a basis for clustering competence centres into distinct types when combined with additional data stemming from case studies, interviews and interim evaluations of centres. As such, the available data reduced complexity for evaluators while at the same time allowing to track different roles and strategies of centres.

However, the monitoring system also revealed major shortcomings and hence options to further increase its usability, in particular for competence centre programmes. The monitoring data under observation put

a strong focus on centres whereas the role of scientific and industry partners who are immediate beneficiaries of these network-type programmes was fairly neglected. This included in particular the role of partners of the centres and the level of engagement of partners, patterns of R&D projects performed with partners and their results, and indications on benefits for participating organisations. Hence, little information concerning key performance dimensions such as knowledge creation and circulation in the participating organisations, but also regarding the internationalisation of their research activities, were obtained by centre monitoring activities.

At a more operational level, the measurement of performance of competence centres was limited to mere counts of publications, patents and PhD/Master theses, and despite the existence of evaluation plans, no comparisons between the performance and the initial, explicit objectives had been made.

Data were further collected in a manner (plain text documents and excel documents) that did not easily allow for incorporation of linkages to external databases. In particular the data collected on publication and patent data and on basic characteristics of the partner organisations were not suitable to be linked with external databases. This limits the use of monitoring data for measuring the long-term impacts of the competence centre programme. While this may have also been a matter of availability of data management software existing at that time, it was yet surprising that programme management did not make efforts to synthesize qualitative information at hand and use it for mutual learning purposes.

TARGETING MONITORING SYSTEMS TO DIFFERENT USER NEEDS

As expressed in the "Platform fteval Evaluation Standards" monitoring data serve the purpose of project controlling in scientific and financial terms and monitoring data should also give evaluators an appropriate insight into the respective project (Platform fteval 2005). Emphasizing performance aspects, monitoring can further be defined as a continuous assessment of key programme functions organised internally by programme management (or a monitoring unit) and carried out on an on-going basis; as such it entails setting up a data collection system for compiling key data on programme activities, participants, interim achievements, and outputs (Dinges 2011).

Hence, when designing monitoring systems of research and innovation programmes, key needs and purposes for monitoring systems should be taken into account (figure 2). In the case of competence centre programmes, four types of customers can be differentiated: programme owners, programme management, centre management and evaluators. For programme owners, financial accountability and cost efficiency of operations are key, whereas for evaluators, monitoring data should at least provide basic data for characterisation of centres/networks and activities performed within these networks. For programme and centre management, monitoring systems should further provide a continuous overview about progress made towards objectives set, and hence allow to steer and fine-tune the development of centres.

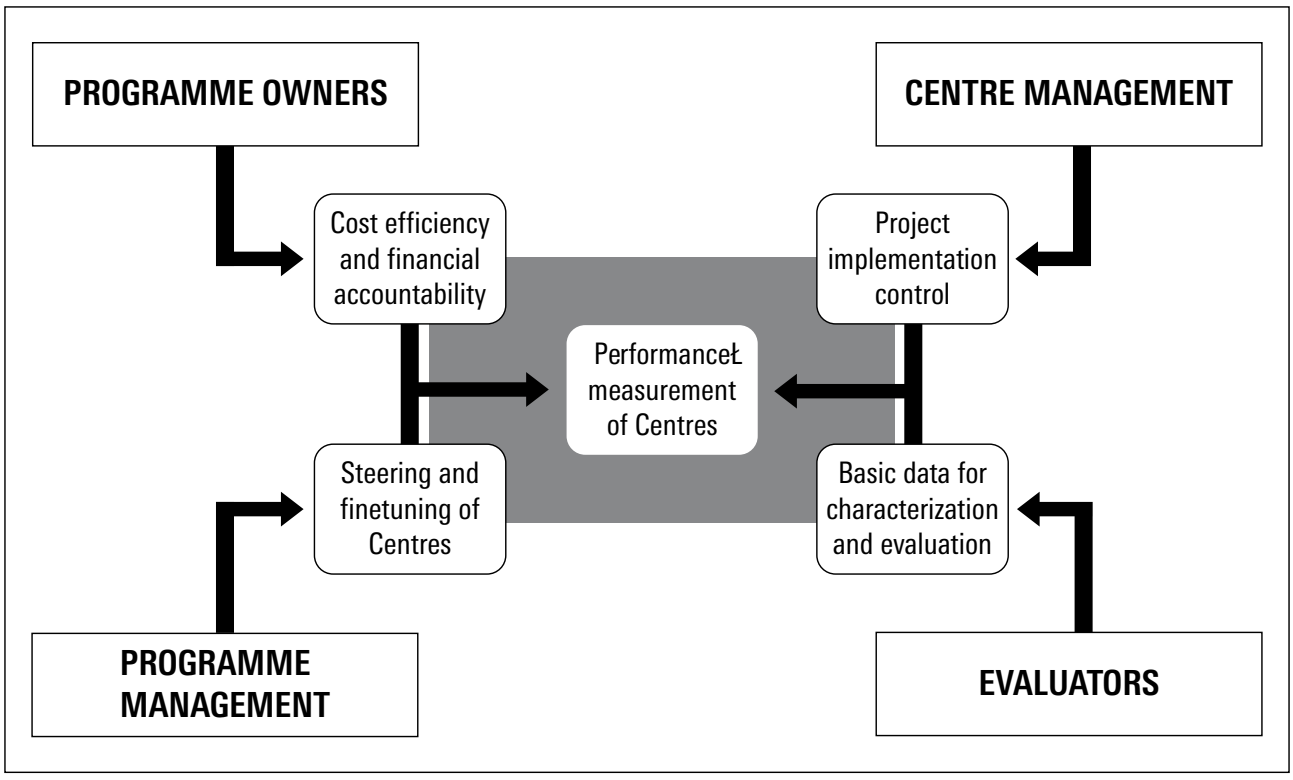


FIGURE 2 - Customers and use of competence centre monitoring data
Source: Own illustration

With the exception of evaluators, all customers of a competence centre have an immediate interest in using monitoring data on a regular basis, as centre management, programme management and programme owners should be able to inform each other and their respective stakeholders about ongoing processes, results achieved, and actions to be set in case of missing objectives.

DESIGN PRINCIPLES OF MONITORING SYSTEMS

It is one of the key advantages of monitoring systems to deliver information in time, while programmes are running. However, The creation of performance monitoring tools should not only be seen as an accountability and implementation tool but primarily as a learning tool to understand the specific structures that emerge and in doing so to improve both the effectiveness and adequacy of implementation of competence centres. When designing performance-monitoring systems, all functions – accountability, supporting implementation and understanding the emergent characteristics of the entities that are funded - must be taken into account.

Consequently, for the immediate function to support the implementation, timeliness (is the monitoring system delivering results when they are needed?), comparability (can the information of individual centres be compared across centres, with similar programmes, other funding mechanisms?), and feasibility (what burden does a monitoring system pose on its constituents) are key considerations for the principle design of monitoring systems.

To support evaluation and learning, the monitoring system should be tailored around the intervention logic of a programme, considering inputs and activities of competence centres, outputs, and outcomes (see e.g. Kavlie and Sleeckx 2011).

As our paper has shown, for competence centre programmes, an indicator framework should be able to track a) actors within the competence centres and b) the portfolio of projects established by competence centres. The involvement of industrial actors and research organisations, each described with key economic and innovation data, allows assessing the degree of heterogeneity of actors and the potential size of the competence centre networks. Both economic and innovation data are critical for econometric analyses to be performed in interim and ex post evaluations. While this provides some basic feature of the competence centre network, only data that combine actors' data and project data allow to analyse network characteristics and development of the networks:

- Project data should therefore include measures such as: the duration and size of projects, the technological/scientific fields covered, the type of research/innovation problem tackled, and the embeddedness of the project within the strategic research portfolio of the competence centres.
- Project partner data should focus on the role of partners within the projects. Key indicators in this respect are the number and share of new participants, new coordinators, new collaborations, frequency of participation of individual organisations, and repetitive cooperations. The indicators are equally relevant for all type of competence centres and thematic areas and can be compared with each other.

As regards outputs, indicators need to show how a programme is progressing towards meeting its objectives. A number of current monitoring systems only provide information on the number of scientific articles, the number of patent applications (see Dinges et al. 2011, Kavlie and Sleeckx 2011). For improving monitoring of outputs from competence centres, competence centre programmes should focus on measures describing progress towards objectives set and detail results of competence centre programmes with regard to various impact/exploitation channels of research (see Martin and Tang 2007). Apart from the knowledge creation dimension, the knowledge circulation dimension/networks and the human capital capacity building dimension seem to be most relevant for network type programmes.

Programme outcomes and impacts may rather be measured in interim and ex post evaluations than in monitoring exercises. Attribution problems and timing are the main reasons why outcomes and impacts are rarely considered in programme monitoring. However, as competence centres frequently have a life time of more than 5 years, reporting systems may also take into account achieved innovation impacts at the level of participants, in particular related to capacity building for different types of product/process innovations, capability to introduce organisational innovations, increased creativity and skills, and economic results achieved through participation in the competence centre network.

For output and outcome measurement, major sources in this regard are periodic project reviews and annual reports to be provided by competence centres. In order to increase usability of this type of data, the use of categorized variables needs to be increased in monitoring systems, whereas for the purpose of evaluations automated text analysis tools need to be developed that allow analysing large sets of fairly unstructured qualitative data in reporting systems.

REFERENCES

Arnold, E., Deuten, J., van Giessel, J-F (2004), An International review of Competence Centre Programmes, Technopolis.

Biegelbauer, P. (2007), Learning from abroad: the Austrian competence centre programme Kplus. Science and Public Policy, 34(9), S. 606-618.

CREST (2008), Report of the CREST Working Group on Industry-Led Competence Centres – Aligning academic / public research with Enterprise and industry needs, http://ec.europa.eu/invest-in-research/pdf/download_en/illc.pdf.

Dinges, M., Poel, M., Sondergaard-Laugesen, N. (2011), Beyond Patents and Publications, Performance Monitoring Indicators for ICT Research in the EU-funded RTD, Joanneum Research, http://www.joanneum.at/uploads/tx_publicationlibrary/Beyond_Patents_and_Publications.pdf.

Edler, J., Berger, M., Dinges, M., Gok, A. „The Practice of Evaluation in Innovation Policy in Europe, Research Evaluation, Research Evaluation, (21),pp. 167–182.

Edler, J., Lo, V., Sheikh, S. (2004), Assessment "Zukunft der Kompetenzzentrenprogramme (K plus und K_ind/net) und Zukunft der Kompetenzzentren". Endbericht. Studie im Auftrag von BMVIT und BMWA.

Kavlie, D., Sleeckx, E. (2011), Report on Monitoring and Evaluation of Competence Research Centres (CRC), IWT Study Report 69, IWT, http://www.iwt.be/sites/default/files/varia/iwt_studie69.pdf.

Lepori B., Barré R., Filliatreau G. (2008), New Perspectives and Challenges for the Design of S&T Indicators, Research Evaluation 17 (1), 33-44.

Martin, B., Tang, P., 2007. The benefits from publicly funded research, SPRU Working Paper Series 161, SPRU - Science and Technology Policy Research, University of Sussex.

OECD (2004): Public-Private Partnerships for Research and Innovation: An Evaluation of the Austrian Experience. Paris.

Ohler, F., Stampfer, M. (1999): Konzept zur Evaluierung der Kplus – Kompetenzzentren: Grundregeln und Herstellung einer Informationsbasis. TiG, Wien.

Platform fteval (2005), Evaluation Standards in Research and Technology Policy, Vienna.

Schibany, A., Dinges, M., Reiner, Ch., Reidl, S., Hofer, R., Marbler, F., Leitner, K-H., Dachs, B., Zahradnik, G., Weber, M., Schartinger, D., Edler, J. (2013), Ex post Evaluierung der Kompetenzzentrenprogramme Kplus und K_ind/K_net, Wien. http://www.bmvit.gv.at/innovation/strukturprogramme/downloadsstruktur/kprogramme__eval__endbericht.pdf.
Schmidmayer, J., Berger, M., Kornberger, B. (2010): Beitrag zur Additionalisitätsmessung bei Unternehmen im Rahmen des Kplus-Programms.

Steyer, F. (2006): Behavioural Additionality in Austria's Kplus competence Centre Programme. In: OECD, Government R&D funding and company behaviour: measuring behavioural additionality, Paris, S. 75-89.

Steyer, F., Gretzmacher, N., Schibany, A. (2007): Beitrag zur Additionalitätsmessung bei Unternehmen im Rahmen des Kplus-Programms, Wien.

TiG (2000): Vorschriften für das Kplus-Berichtwesen, Wien.

AUTHORS

MICHAEL DINGES
Innovation Systems Department
E: michael.dinges@ait.ac.at

WOLFGANG POLT
Joanneum Research, POLICIES
E: wolfgang.polt@joanneum.at

K. MATTHIAS WEBER
AIT Austrian Institute of Technology, Innovation Systems Department
E: matthias.weber@ait.ac.at

EVALUATION OF CLUSTERS, NETWORKS AND CLUSTER POLICIES – CHALLENGES AND IMPLEMENTATION

A PRACTICE-ORIENTED APPROACH FOR IMPACT ASSESSMENT THROUGH EVALUATION CONTRIBUTING TO A FURTHER DEVELOPMENT OF CLUSTERS AND CLUSTER POLICIES

SONJA KIND, GERD MEIER ZU KÖCKER

INTRODUCTION

**THE EVALUATION OF CLUSTERS AND CLUSTER POLICIES RAISES CERTAIN CHALLENGES –
EFFECTS CANNOT EASILY BE DETERMINED.**

Cluster policy has been enjoying high popularity for many years as it has become an increasingly important instrument of modern business and trade promotion. In order to make sure that the EU will henceforward allocate resources from the Structural Funds, the member states, and the regions, respectively have to present innovation strategies for an intelligent specialisation. Thus, the role of clusters in Europe will become ever more important in the future as they are considered as the central element of Smart Specialisation.

Expectations regarding clusters are high. They are supposed to enhance the competitiveness of regions or locations as well as that of the companies and research institutes operating in these areas, what usually leads to job creation and prosperity in the medium term.

The evidence, whether cluster policy actions could effectively contribute to the realisation of these objectives, is moving further into the focus of political decision-makers and programme initiators of cluster policy. Relevant aspects hereby include on the one hand the legitimacy of fiscal expenditure and on the other hand, the possibility to intervene in relevant processes in a proactive way.

Given the complexity of clusters, a systematic impact assessment seems to be utterly impossible. In theory, the evaluation of a political intervention aims however at the assessment of the additional effects directly caused by a policy initiative.

The fact that there are practically no reference situations with an absence of policy action effects leads to particular difficulties – a real "untreated" control group in this respect will not be available when it comes to cluster evaluation.

Each cluster is unique due to its own history, industry and geographic location. There is also a variety of incontrollable factors of influence, for instance further political interventions, such as other incentive programmes or the general (macro-)economic development. Moreover, impacts often become visible only after a considerable delay.

In order to record the effects at company, cluster and regional level, and thus to meet the high demands of a valid evaluation, in theory, clusters would actually have to be analysed by using an extensive mix of quantitative and qualitative methods. The methods available are numerous and range from case studies to network and econometric analyses. In practise, such an approach is failing not only due to the related costs and time spent, but first and foremost also due to the lack of available data and their insufficient quality. A query of sensitive company data, for instance, will mostly lead to very low response rates and would thus make an analysis impossible.

When evaluating clusters, impacts at company level as well as the effects related to the cluster itself and the respective region as a whole are of particular interest. The instruments that can possibly be used for an impact assessment include, for example, surveys or the above mentioned econometric approaches.

In the context of several evaluation projects, the iit – Institute for Innovation and Technology (Berlin) has developed an applied practice-oriented strategy for the evaluation of clusters and cluster policy. The approach takes all levels into account that are relevant for the clusters: the cluster policy in place, the cluster management organisation as well as the cluster actors. Furthermore, the medium- to long-term effects are measured by using a methodological mixture based on surveys, interviews, workshops and benchmarking.

EVALUATION OF CLUSTERS AND CLUSTER POLICY – THE IIT APPROACH

THE EVALUATION AIMS TO IMPROVE THE PERFORMANCE, EFFECTIVENESS AND SUSTAINABILITY OF CLUSTERS.

In general, the evaluation contributes to a process of mutual learning and knowledge exchange at the relevant actor levels, namely cluster policy, cluster initiative and cluster management organisation. On the basis of the evaluation results, the **performance**, **effectiveness** and **sustainability** of ongoing projects could be improved

- Performance – with regard to the cluster management organisation
- Effectiveness – referring to the cluster policy and the cluster management process
- Sustainability – with regard to the cluster management process

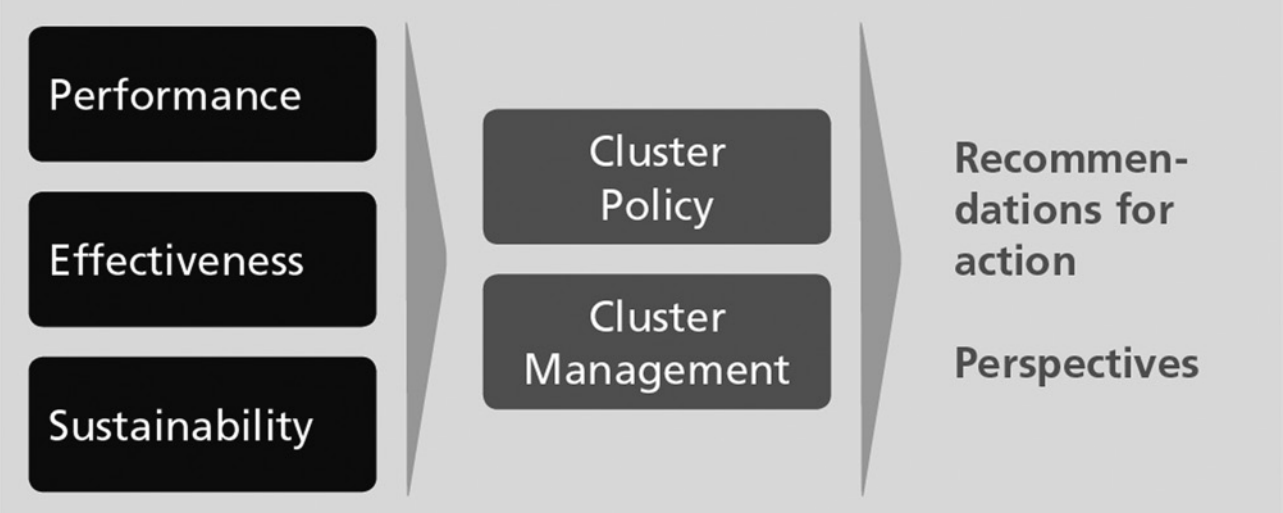


FIGURE 1 - Levels of analysis of the evaluation

THE EVALUATION APPROACH INCLUDES FOUR MAJOR PROCESS PHASES:

At the beginning, the objectives, the subject and criteria of evaluation will be determined. Herby, different cluster stakeholders are involved. This way, transparency is guaranteed which helps to achieve a result that will later be accepted by all parties involved.

FIGURE 2 - Evaluation process

In a second step, the required data will be collected and analysed. Thirdly, the results will be discussed and reflected and the prospects for a further development of the cluster policy and the cluster itself will be developed in cooperation with the involved stakeholders. As a final point, the recommendations for action will be recorded together with the other results by means of a final report.

EVALUATION OF THE EFFECTS AND PERFORMANCE

For the evaluation of the effectiveness, the focus is on the following questions:

- How have competitiveness and innovative capability of the cluster actors developed in recent years?
- Which contributions can be attributed to the cluster policy and the cluster management?

When assessing the impact, it is normally assumed that the input (e.g. provided resources) correlates with the performances shown under the implemented measure (output) as well as with the results intended to be achieved by the target group of the policy measure (Outcome). The results of an intervention that are not occurring in the target group, but

in its environment are termed 'impact'. What is to be understood by the terms input, output, outcome and impact in connection with a cluster and network evaluation will be explained in the box below:

INPUT, OUTPUT, OUTCOME, IMPACT – IN CONNECTION WITH THE EVALUATION OF CLUSTERS AND NETWORKS

- The term 'input' primarily refers to financial, human or other resources being invested. In most cases, these include influencing factors, such as types of costs (budgets for personnel or material resources) or personnel qualification schemes (influencable by training programmes). The main input factors are the competences of the personnel being active in the cluster initiatives and cluster management organisations as well as the available budgets.

- The term 'output' describes all performances, such as activities, publications and particularly services being directly produced by the cluster initiative, including brochures, workshops, coaching, counselling interviews, events etc.

- The term 'outcome' includes the results intended to be achieved by means of interventions/activities of a cluster programme, such as changes in attitudes or behaviour of the target group members or benefits for the target groups. Unintended results described by the target group members do not fall within that definition of an 'outcome'. The target group of the activities of cluster management organisations primarily consists of companies and research institutes organised in the respective cluster initiatives.

'Impacts' are defined as the results achieved by a cluster programme being effective beyond the target group itself and may not be influenced by the cluster management, such as positive macroeconomic effects in a specific region due to a revenue growth or headcount increase. Normally, a description of direct causal efficacies regarding the performances of the subject of evaluation and the impact is not possible (also due to the often occurring timing differences between input, output and impact).

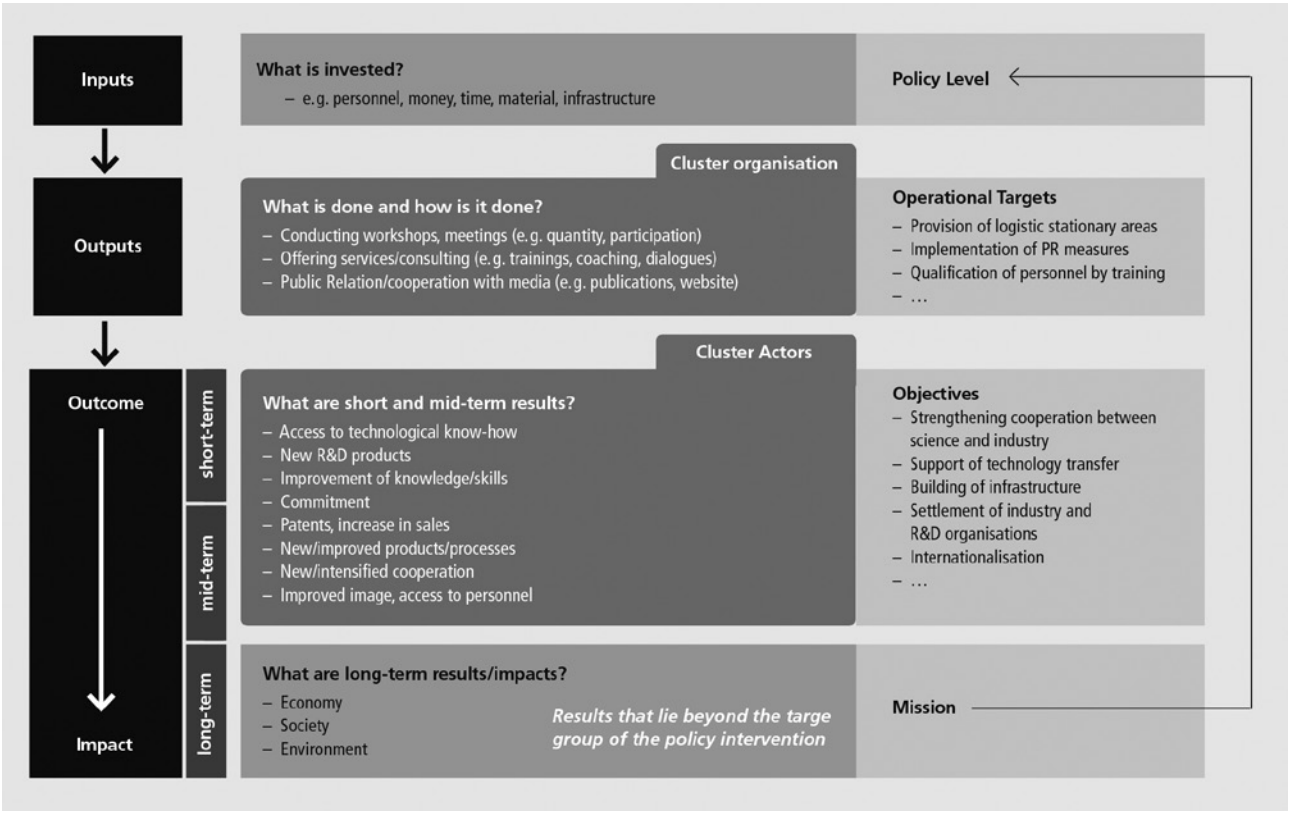


FIGURE 3 - Input, output, outcome, impact – overview

INDICATORS AND METHODS FOR THE CLUSTER AND NETWORK EVALUATION

The output, outcome and impact, i.e. both, short-term and also medium- and long-term results are measured by using a methodological mixture including surveys, interviews, workshops and benchmarking.

As indicated in the figure below, the evaluation system comprises a set of indicators for each of the relevant levels, namely cluster policy, cluster management and cluster actors.

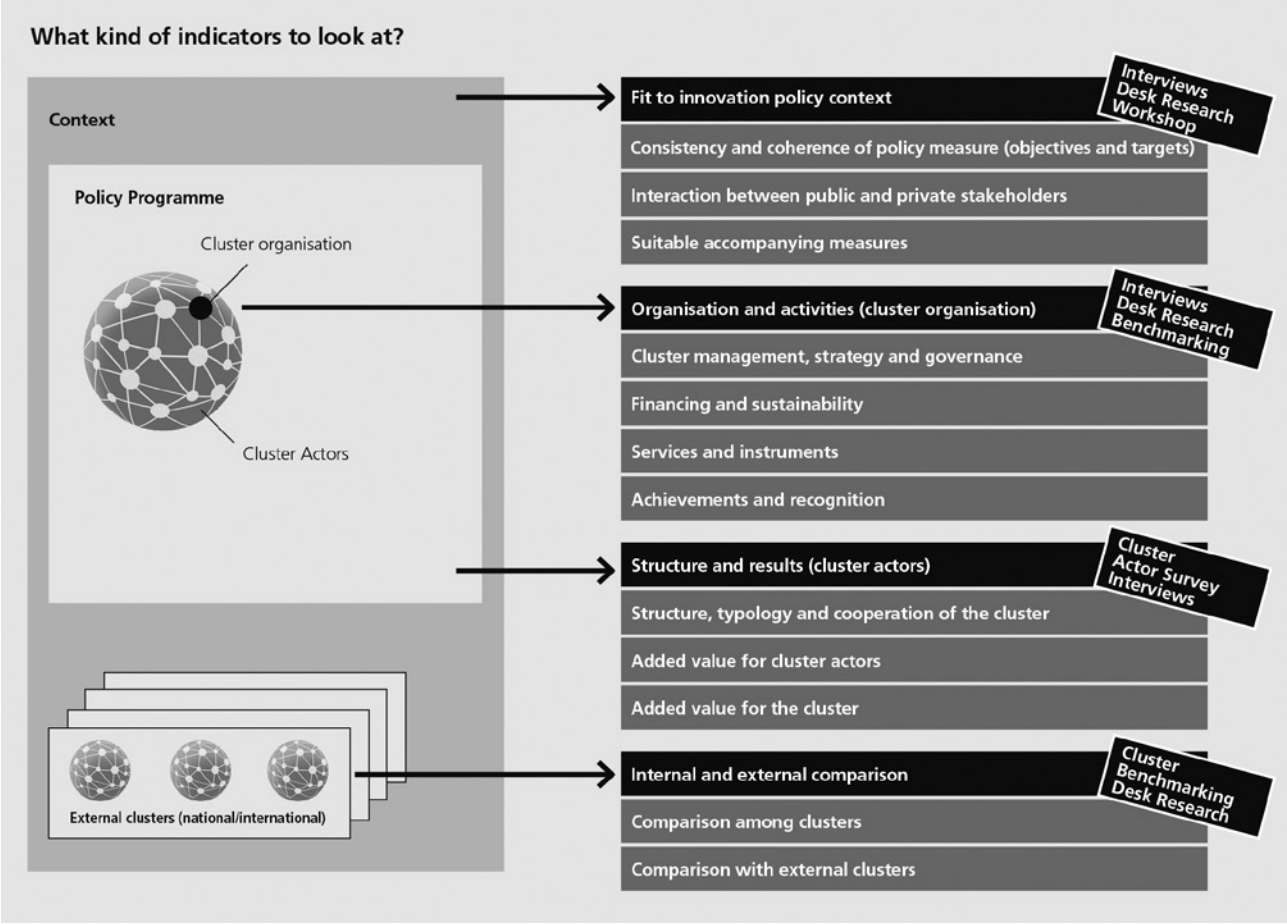


FIGURE 4 - Which thematic levels of indicators should be considered when evaluating clusters and networks?

CLUSTER POLICY LEVEL

With the instrument of cluster funding, cluster policy forms the overarching, constitutive element, whereas its practical implementation is carried out by the cluster management organisations and the actors in the cluster initiatives, respectively. Cluster policy is analysed with reference to its economic and innovation political context and is evaluated in view of numerous and often not influenceable interactions.

Besides general economic and cyclical economic factors of influence, a political measure is always in interaction with other political interventions. Hereby, a differentiation must be made between measures which directly support the own policy, and which have partly been initiated by the respective policy-maker him-/herself, such as specifically introduced supporting strategies for the qualification of the regional personnel or those interventions that are directly influencing local companies, institutions and research institutes.

THE EVALUATION OF CLUSTER POLICY IS FOCUSED ON THREE LEVELS OF ANALYSIS

1) Interaction and organisation of cluster policy at policy level

At this level, it is analysed in which way cluster policy is organised in terms of organisational and institutional aspects, and whether these structures appear suitable for a practical policy implementation. It is then determined whether it is necessary to intensify the actors'/institutions' involvement as well as the coordination between them.

2) Consistency of the objectives of cluster policy and their embedding in the context of economic aspects and innovation policy

This analytical unit examines the question of whether the objectives of cluster policy are congruent with the objectives pursued by the cluster initiatives. After a review of the individual objectives, it needs to be verified whether the objectives of the cluster initiative are inconsistent with the overall objective. In a further step, it will be analysed whether

the followed cluster policy is compatible with other political interventions regarding economic and innovation policy at regional, national and international level.

3) Strategy and prospective orientation of cluster policy

This analytical unit primarily aims to provide a general understanding of the "history" and the initial situation of the cluster initiative compared to its current situation. An important aspect of this issue presents the review of the currently ongoing monitoring and evaluation activities in the individual cluster initiatives as well as the way in which the exchange of information, and thus the process of mutual learning between the cluster managers and personnel in charge, is organised and carried out. The results of the overall evaluation are finally merged and if necessary, adjustment requirements are derived.

CLUSTER MANAGEMENT LEVEL

In the past years, it has been confirmed that the success of clusters does not only depend on the provision of an infrastructure, positive framework conditions and the actors' potentials, but also and specifically on the availability of an efficient cluster management organisation. The cluster management initiates and coordinates the common activities of the cluster initiative, and thus considerably adds to a positive development of the cluster.

At the cluster management level, the analysis is based on the following priorities: In a first step, similarly to the approach at cluster policy level, the topics for interaction as well as the rolls of the cluster management and the respective members will be examined. Furthermore, questions are raised regarding the objectives of the specific cluster strategy and about the realistic capabilities for their achievement. This does also include the question of strategies and whether the achievement of the pursued objectives is regularly measured by means of evaluations or a monitoring process:

- Interactions between the cluster management organisation, cluster members and representatives of cluster policy;
- Consistency of the objectives of the cluster initiative;
- Strategy, monitoring and prospective orientation of the cluster.

One option for analysing the efficiency of the cluster management is to perform benchmarking. Benchmarking is used to examine the potentials of cluster organisations in five different dimensions related to a reference portfolio of more than 250 cluster management organisations that have already been benchmarked. In order to avoid a comparison of "apples and oranges", each cluster management organisation is attributed to a suitable reference portfolio (e.g. comparison of a cluster management organisation of an IT cluster with other cluster initiatives operating in the same technology field).

Consequently, cluster management organisations may not only be evaluated on the basis of a national comparison, but also compared to international standards. The analysis focuses specifically on 33 indicators in total, which are attributed to the following five dimensions:

1. Structure of the cluster organisation and their integration within the cluster initiatives;
2. Management;
3. Financing;
4. Activities and services offered by the cluster organisation;
5. Effects of cluster management on the development of the cluster.

The results of the benchmarking analysis are presented for each cluster organisation in a comprehensive report. They do not only provide information to the management of the respective cluster organisation about potential areas for improvement. The benchmarking with other cluster organisations does also imply the possibility to learn from the reference clusters and to integrate best practice strategies into the own work. This approach for measuring the efficiency of cluster organisations does also constitute the basis for the quality label for cluster organisations developed by the European Cluster Excellence Initiative (ECEI) on the basis of significant input provided by VDI/VDE Innovation + Technik GmbH and the iit. Since 2012 already, cluster organisations that had performed particularly efficiently in the sense of an excellent management have received a quality label (gold label) after a thorough examination by certified experts.

CLUSTER ACTORS' LEVEL

The cluster actors should essentially benefit from the cluster policy measures. The measurement of effects at this level is particularly interesting, but at the same time, it is also particularly difficult due to the above described challenges.

When analysing the impact, it must be taken into consideration that the efficiency of clusters and their actors does not only depend on the capability of interaction and innovation of individual actors or on the given framework conditions. It is rather influenced by the commitment of the respective cluster management organisation. That is why at the cluster actors' level, the analysis is not only focused on the effects and additional values achieved by the cluster actors, but rather on the question whether the service portfolio offered by the cluster management organisation is considered adequate and effective from the point of view of the involved actors.

Hereby, two investigative approaches can be recommended:

1) Satisfaction with and adequacy of the service portfolio offered by the cluster organisation as well as the effects and benefits arising for all members (member satisfaction survey)

The cluster actors are asked to answer questions regarding their satisfaction with the services offered by the cluster management organisation as well as about noticeable additional values. The focus hereby is set on the following question:

- Question about the members' characterisation;
- Questions regarding the members' activities and structure;
- Questions regarding the services offered by the cluster management organisation;
- Questions about the major areas of cooperation;
- Questions regarding the effects of an active involvement in the cluster initiative.

2) Effects and benefits for companies (Cluster Impact Analysis)

The effects achieved by the companies that are actively involved in a cluster – in particular SMEs – are of special interest. It is assumed that companies, especially SMEs, that are organised in a cluster initiative do benefit from an enhancement of their efficiency and competitiveness.

Compared to larger firms, SMEs are usually lacking in resources of various categories (e.g. capital and human resources, qualification schemes amongst others). In clusters, however, SMEs should be able to compensate for their individual deficiencies through the conclusion of strategic partnerships and to generate shared or individual benefits. Companies along the value chain are expected to enter into cooperation with other stakeholders which will finally lead to the development of potential synergies, such as the establishment of new customer relationships or a facilitated access to distribution channels. The special opportunity for enterprises involved in cluster activities is the possibility on the one hand, to focus on their core competencies (specialisation) and on the other hand the potential increase of their limited resources thanks to an integration into a complete system.

The Cluster Impact Analysis examines the following questions:

- To what extent do companies benefit from an active involvement in networks and clusters?
- Which fields can be identified where the positive effects for the networking companies are most apparent?
- Can the objectives that are pursued in line with the entrepreneurial commitment for cluster activities finally be achieved?

For further information about the Cluster Impact Analysis, please use the following link:

http://www.iit-berlin.de/veroeffentlichungen/cluster-impact-analysis-the-real-cluster-case (a German version is available on request).

CONCLUSION

Clusters are individual organisations and should always be evaluated under consideration of their specific business contexts. Our experience with the evaluation approach described in this article shows that this model provides a sufficiently flexible framework. Thus, it allows for an individually tailored proceeding which is able to meet the specific requirements of different contracting parties and clusters. The evaluation model comprises various methodical approaches, and consequently includes a comprehensive set of indicators from which the adequate ones can be selected.

In brief and above all, the particularity of this approach is characterized by the following facts:

- Specific characteristics of the clusters and networks (age, intensity of commitment within the cluster) are taken into account.
- Due to comprehensive practicability tests and a consultation process, the applied indicators enjoy broad acceptance among political decision-makers and cluster actors.
- The impact analysis is not based on the collection of sensitive business data, but on an entrepreneurial assessment of potentially achieved results attributed to predefined success categories;
- This type of analysis ensures an independency from the availability of statistical data and predefined sectors and regions, respectively. Technologies and activity fields do often not correspond to the sectors that are displayed according to the industrial classification. Moreover, clusters and networks are normally difficult to reconcile with statistically covered territorial units.
- The analysis focuses on data that are comparably easy to collect and to evaluate.

This investigation method does not require complex calculations that would be perceived as 'black box' - leading to easily comprehensible results.

AUTHORS

SONJA KIND
iit – Institute for Innovation and Technology
E: kind@iit-berlin.de

GERD MEYER ZU KÖCKER
iit – Institute for Innovation and Technology
E: mzk@iit-berlin.de

MEASURING PRODUCT INNOVATION AND INNOVATIVE CAPACITY: NEW INDICATORS TO EVALUATE RESEARCH PROGRAMMES

CHRISTIANE KERLEN, ERNST A. HARTMANN

INTRODUCTION

Measuring innovation is one of the main tasks in evaluating research, technology and innovation programmes. Quite a few indicators have been developed to accomplish this task and to allow for an internationally comparative perspective (Jaedecke et al. 2009, Luoma et al. 2011, Peters 2012). Nonetheless the overall puzzle of measuring innovation is far from being solved. There is still work to be done in developing those instruments that capture innovation and innovative capacity best.

A widely accepted typology of innovation categorises four different types (Tiwari 2008, Oslo Manual 2005): Product innovation refers to the introduction of a product or service that is new or significantly improved with respect to its characteristics or intended uses. Process innovation is the implementation of a new or significantly improved production or delivery method. Organisational innovation refers to the implementation of new organisational structures or processes in the firm's business practices, workplace organisation or external relations. Marketing innovation describes the implementation of a new marketing method involving significant changes in product design or packaging, product placement, product promotion or pricing.

This article will mainly address the first and the third type of innovation with a twofold approach. Firstly, it will focus on measuring steps of product innovation during research and development. If there is a long time span between basic research, research and development, and market-ready products, it becomes necessary to measure 'in-between' steps along this research, development, and innovation process. The 'Technology Readiness Level (TRL)' approach is used as a metric for these intermediate steps. Secondly, it will focus on measuring innovative capacity: Regarding these long time spans, it makes sense to measure not only impacts on innovations, but also on innovative capacity. Innovative capacity relates to companies´ ability to produce innovation (Cedefop 2012; Cohen/Levinthal 1990). One core aspect of innovative capacity is the innovation-conduciveness of organisational structures within the companies. Thus, an analysis of innovative capacity will also shed light on relations between organisational and product innovations.

BACKGROUND

The ideas outlined in this paper have recently been tested in an evaluation of a large German innovation and technology programme. From January 2012 to March 2013, the German Aviation Research Programme has been evaluated by the VDI/VDE-IT's Institute for Innovation and Technology (iit) in collaboration with other experts1 on behalf of the German Federal Ministry for Economics and Technology. A principle task of this project was to develop an indicator-based evaluation system for the aviation research programme. This system formed the basis for an ex-post evaluation of the third aviation research programme LuFo III (that ran between 2003-07), a formative evaluation of the fourth research programme LuFo IV still in progress during the evaluation (since 2006, 150 Mio € p.a.) as well as an ex-ante evaluation of the fifth programme LuFo V that will run from 2014. The aim of the evaluation was to analyse the effectiveness, efficiency and efficacy of the programme – in accordance with the objectives of the German Federal Government – and perform an assessment in order to make recommendations for the management of the programme and the organisation of possible follow-up programmes. One specific task within this evaluation was to analyse the aviation industry and its needs to make recommendations on future research topics.

Due to the specific characteristics of the aviation industry, the evaluation team was confronted with a major problem: very long time spans until first monetary results should be visible. As shown in Figure 1 for LuFo III, a programme that ran from 2003 to 2007, in only 19 companies monetary effects had been realized at the time of the survey. There are some companies that expect monetary effects to occur 30 years later than the research and development project has been conducted. Altogether the time frames until first monetary results are expected are somewhat shorter for LuFo IV (since 2006). This can be explained by the fact that the selection committee was aware of the problem and tried to choose projects with shorter horizons of completion. Still timespans reach up to 25 years.

1 The evaluation was supported by experts in the aviation sector, an expert in macro-economics and an evaluation expert, namely Dr Christiane Kerlen.

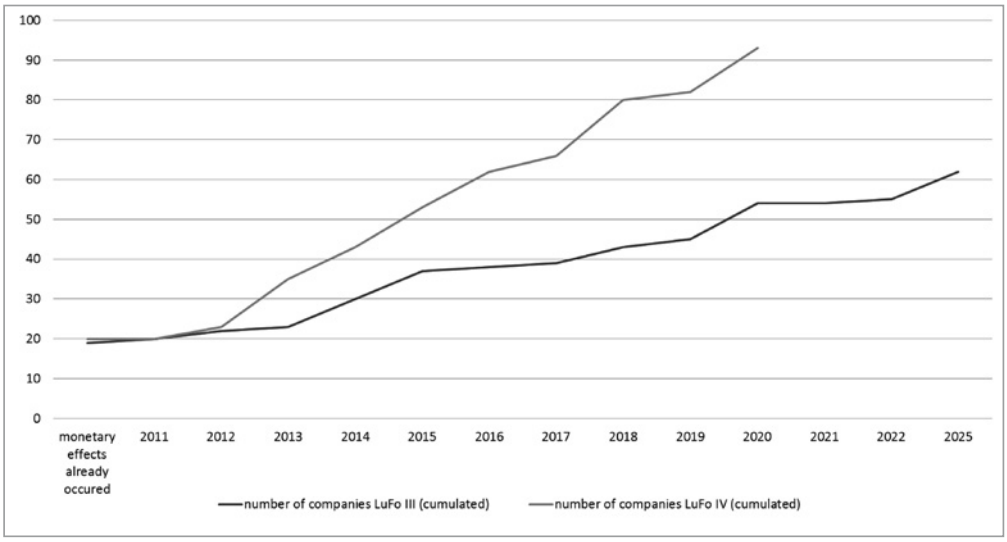


FIGURE 1 - Time spans for monetary effects to occur
LuFo III UN, n=62; LuFo IV UN, n=93

METHODOLOGICAL FRAMEWORK: INDICATORS FOR EARLY MEASUREMENT OF OUTPUTS AND OUTCOMES

To measure 'in-between' steps along this long way from research, development, and innovation to monetary outcome, two indicators were established to bridge the gap: One indicator to show the accomplished level of technology development, the other to show the effects on innovation capacity within the participating companies.

TECHNOLOGY READINESS LEVEL (TRL)

One way of assessing the development of a technology is to assess the technology readiness level. Using Technology Readiness Level (TRL) allows assessing the level of technology matureness. The levels span the earliest stages of scientific investigation to the successful use in a system/ implementation in the market. Decision making on investments in research and development activities can be assisted by this concept. TRL are focussing on the technological matureness of the (future) product. Therefore they are just one of a set of criteria for investment decisions. For example, complementary concepts exist, e.g. of Integration Readiness Level (IRL) or Integration Maturity Level (IML), that focus on the integration into a system (e.g. an airplane), or Manufacturing Readiness Level (MRL), focusing on cost effective production (DIN NL/AST/DRL 2008).

The concept of TRL was developed in the beginning of the 1980ies by NASA[2]. They are being applied by the US administration as well as military sourcing organisations and aerospace agencies. They are also being used in civil areas and are meant to be a technology independent concept that might need adjustments by experts in certain technology fields (DIN NL/AST/DRL 2008; Graettinger et al. 2002). An International Standards Organization (ISO) definition of the Technology Readiness Levels (TRL) and their criteria of assessment has been accepted in spring 2013 (Bilbro 2010, 2011), making the concept even more applicable in other fields.

One can distinguish between TRL in a narrow sense that are valid for airplane specific hardware. An ESA/NASA-definition of Technology Readiness Levels can be used here. For other development projects a more general description can be used which was derived by the US-DoD. It allows the assessment of research and development (R&D) projects that do not focus on airplane specific hardware (i.e. development of software-tools, production technologies, etc.) (DRD/DDR&E 2009, ASD (R&E) 2011). An adjustment for a specific technology field generally has to be made. Therefore there are different definitions for different fields, for example for biomedical TRLs (DRD/DDR&E 2009, Appendix E).

Broadly speaking, publicly funded programmes are intended to overcome market failures. For research programmes this failure can be identified as a high risk for a single company to invest in research and development (R&D) because the outcome of a research project is highly uncertain. Sometimes failures are necessary to find the right way, and in general ways to exploitation are intricate. From a macroeconomic perspective it is desirable that private companies invest in research and development. Only a few innovations might be the ones that can keep a whole industry competitive. A publicly funded research programme can therefore reduce the risk for a single company, because the investment the company has to make is reduced by the funding.

With regard to TRL only low levels of technology readiness can be funded in research, development and innovation programmes if this risk is to be reduced, generally TRL lower than 6. To use TRL as a metric for measuring the progress within a single project, one may ask for the TRL at the beginning of the funded project, the initially targeted TRL, and the actual outcome. Additionally, it can be asked what TRL has been achieved at the time the survey is taking place.

2 National Aeronautics and Space Administration

TECHNOLOGY READINESS LEVEL TRL	GENERAL DESCRIPTION (US DOD DEFINITION)
TRL 1	Basic principles observed and reported
TRL 2	Technology concept and/or application formulated
TRL 3	Analytical and experimental critical function and/or characteristic proof of concept
TRL 4	Component and/or breadboard validation in a laboratory environment
TRL 5	Component and/or breadboard validation in a relevant environment
TRL 6	System/subsystem model or prototype demonstration in a relevant environment
TRL 7	System prototype demonstration in an operational environment
TRL 8	Actual system completed and qualified through test and demonstration
TRL 9	Actual system proven through successful mission operations

DoD-Definitions, DRD/DDR&E 2009
TABLE 1 - TRL definitions

Desk research, expert interviews and pre-tests for the evaluation of the German Aviation Research Programme showed that TRL assessments are widely used within the aviation industry. All project leaders therefore were acquainted with the overall concept and the specific definitions (ASD (R&E) 2011; DIN NL/AST/DRL 2008).

INNOVATIVE CAPACITY

Innovative capacity often is an explicit objective of research programmes. Innovative capacity is a core prerequisite for actual innovation, and is attracting more and more awareness in discussions relating to innovation analysis and measurement. Assumptions regarding the underlying causal relationships are shown in Figure 2. It is assumed that public funding can influence individual learning by developing new educational formats (formal or non-formal learning, i.e. further education with or without a degree), or because informal learning conditions are changed by a programme, e.g. by introducing new, more intellectually demanding tasks and operating procedures, or new organisational structures and processes. This individual learning also contributes to organisational learning processes. On the other hand, public funding can influence organisational learning directly. Organisational learning can be conceptualized as building up on the organisation`s intellectual capital which encompasses the three dimensions of human capital, structural capital, and relational capital (Alwert 2005; see next paragraph).

Impacts of public funding on organisational learning may include the stimulation of new forms of formal or informal learning in the companies (human capital), new organisational processes and structures (structural capital), or new external relations to customers, suppliers, competitors, educational and research institutions (relational capital). Intellectual capital in this sense can be regarded as the basis of innovative capacity.

Individual and organisational learning help to be more innovative and to produce more innovations, especially product, product and organisational innovation. Conversely, being innovative and innovation also have a positive 'backwards' impact on individual and organisational learning, because innovation itself provides a broad range of learning opportuni-

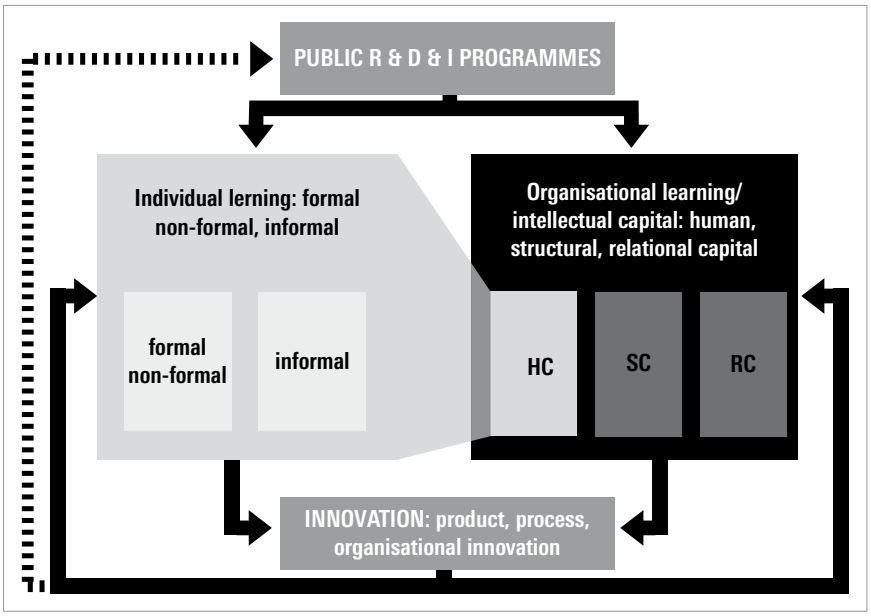


FIGURE 2 - Assumptions on causal relationships between public R&D&I programmes and innovation

ties. If the programme and its impacts are evaluated, taking into account the impact dimensions as relating to innovative capacity, this may have an impact on the future design of research, development and innovation programmes.

As mentioned before, the intellectual capital of an organisation can be regarded as the basis of its innovative capacity. It encompasses three dimensions: Human capital describes the knowledge and skills of employees, the human resources development, and provision of continuing vocational education by employer. Structural capital refers to R&D structures, departments, technological equipment, R&D processes, communication and cooperation between R&D and production, learning-intensive and innovation-conducive organisational structure, and learning and innovation-oriented corporation culture. Relational capital encompasses relations to customers, suppliers, research institutions, educational, services, and the general public.

There are strong relationships between core aspects of innovative capacity and actual innovation performance. This is illustrated in Figure 3. In the upper graphs, correlations can be seen between workplace learning and task complexity, two aspects of informal learning as an aspect of human capital formation, and innovation performance on a national level. The lower graphs show correlations between participation in adult learning (human capital) and type of work organisation (structural capital), and national innovation performance. (CEDEFOP 2012).

Human capital	Structural capital	Relational capital
• Knowledge and skills of employees • Formal (degree) • Non-formal (further education without degree) • Informal (learning by doing) • Human resources development, provision of continuing vocational education by employer	• R&D structures • departments • Technological equipment • R&D processes • Within R&D • Communication and cooperations between R&D and production • Learning-intensive and innovation-conducive organisational structure • Learning and innovation-oriented corporation culture	• Relations to • customers • suppliers • Research institutions • educational services • General public • Image and brand

TABLE 2 - Three dimensions of intellectual capacity/innovative capacity

APPLICABILITY IN AN EVALUATION

Both of the indicators have been tested and applied in the evaluation of the German Aviation Research Programme. Applicability in this context shall be discussed in the next sections.

USING TRL – RESULTS OF THE GERMAN AVIATION RESEARCH PROGRAMME

For the evaluation of the German Research Aviation Programme a differentiation between hardware and software related TRL was made using a filter question. Based on this definition, the project managers were asked in the online survey which TRL had been achieved before starting the project, which TRL they wanted to achieve as a result of the funded project and – for projects that continued working on the topics – which TRL they had reached at the time of the survey. The results of the ex-post evaluation (LuFo III) are shown in Figure 4.

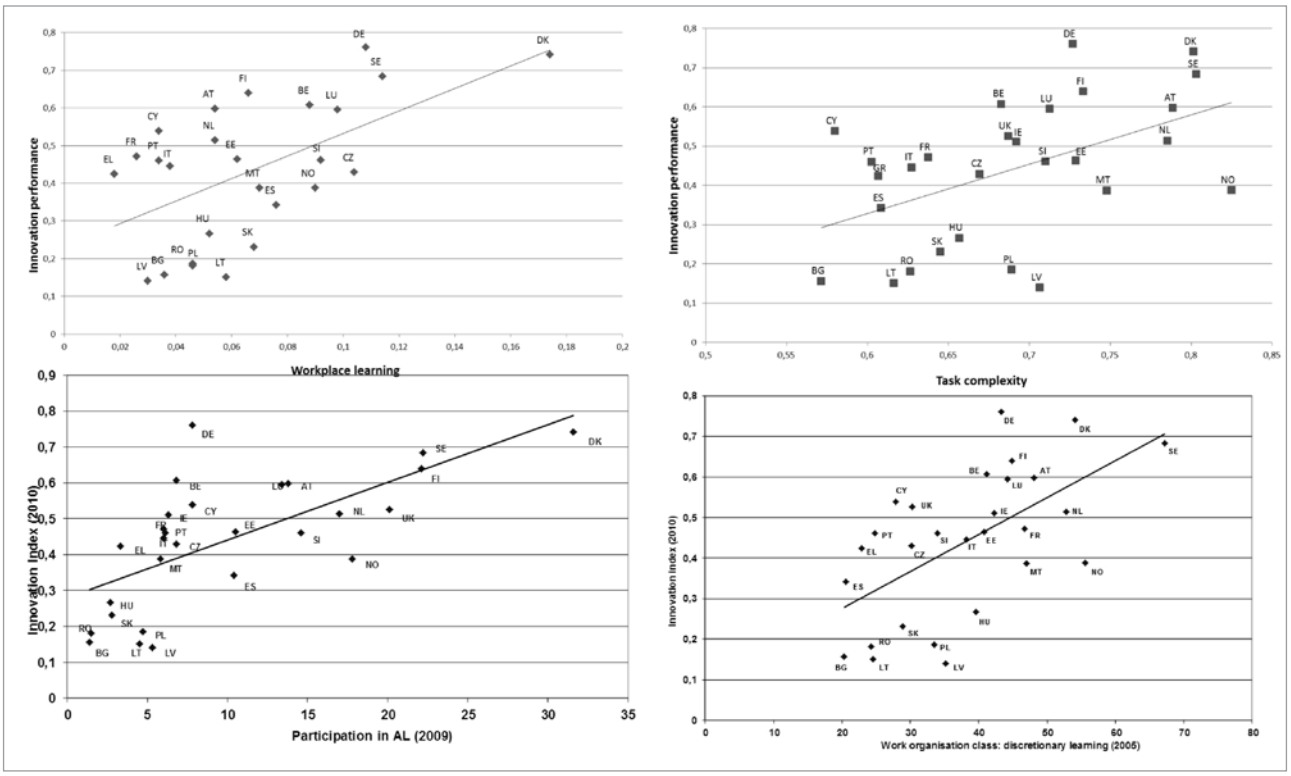


FIGURE 3 - Relationships between core aspects of innovative capacity and innovation

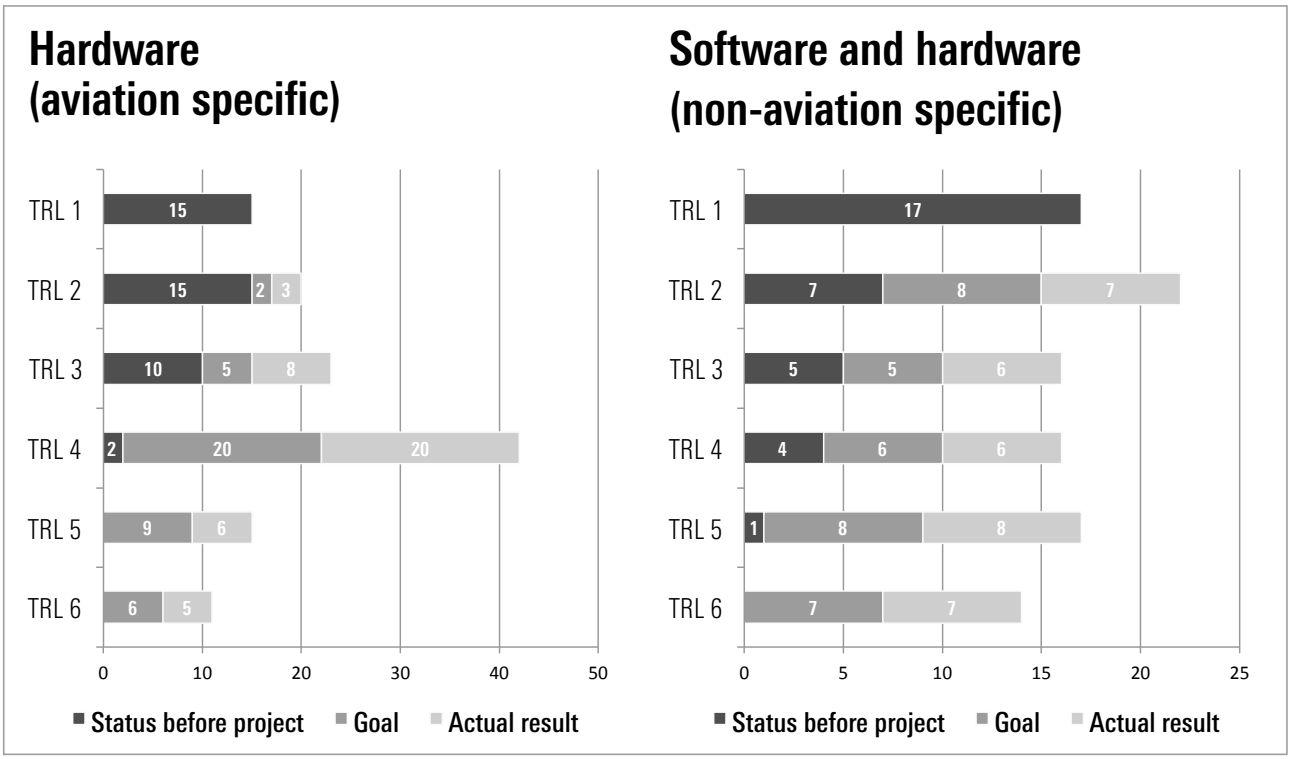


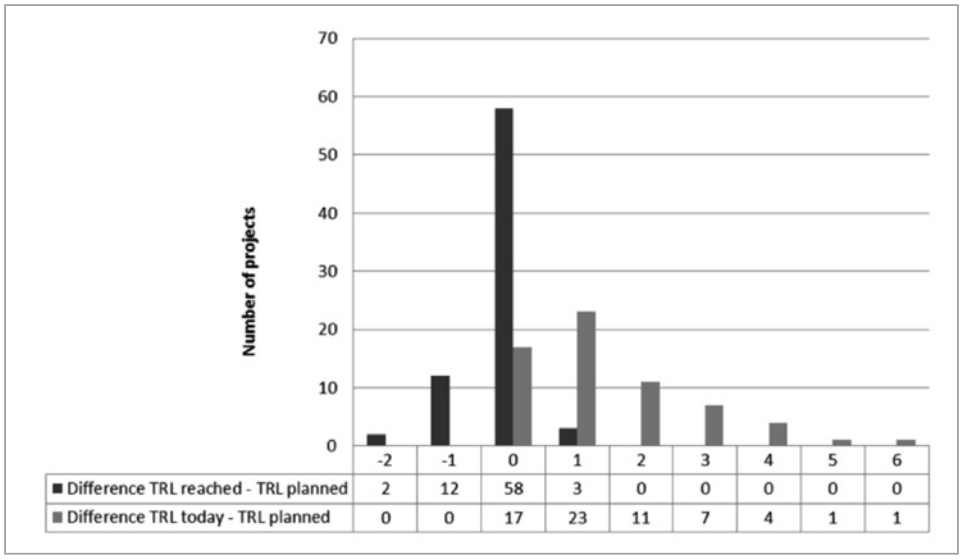
FIGURE 4 - Status before project, goal and actual result in TRL Projects conducted by private companies with the funding of LuFo III (only aviation specific hardware, n=44; only non-aviation specific hardware and software, n=34)

For aviation specific hardware, most of the projects started with a TRL of 1 or 2, some with 3 or more. An advancement of two levels was targeted at in most of the projects; quite a few projects were only aiming at bridging one level. In some cases an advancement of more than two levels was set as target. For non-aviation specific hardware and software, the majority of projects started on TRL 1. Here an advancement of three levels was observable. On average this means that companies assume slightly shorter development times for non-aviation specific technologies. Overall it can be seen that LuFo-funding aims at areas with a high research risk and long timescales.

Based on the data shown in Figure 4 an analysis of the differences between original project goals and realised TRL was made (see Figure 5).

Aviation-specific and non-aviation-specific developments are not separated in this figure. One can see that for a small amount of projects the originally targeted TRL was outrun (4 percent). For 77.3 percent of all projects the targeted TRL was achieved and in 18.7 percent of the cases the targeted TRL was not met. Even though the majority of the projects reached the targeted TRL, this result also illustrates the uncertainty of research activities. This holds even more considering the fact that projects successfully reaching a TRL of 4 or 5 will not necessarily be reaching market readiness. Independently of meeting the targets the vast majority of projects in companies (94.7 percent) stated a technological progress in the field they had been researching.

Regarding the longer term effects of the programme it is interesting to see if the publicly funded projects were further pursued in the companies. For this reason the projects were asked at what TRL the projects were at the time of the survey. 73.4 percent answered (figure C, orange) that original technology is on a higher readiness level today. The high number of projects indicating development progress of two levels or more indicates that technologies are developed consequently towards a readiness of production. 19 companies stated that their technology has reached a readiness level higher than 6 which is close to exploitation.



	-2	-1	0	1	2	3	4	5	6
Difference TRL reached - TRL planned	2	12	58	3	0	0	0	0	0
Difference TRL today - TRL planned	0	0	17	23	11	7	4	1	1

FIGURE 5 - Differences between planned TRL at project end and actual result and between TRL reached today and planned TRL at project end

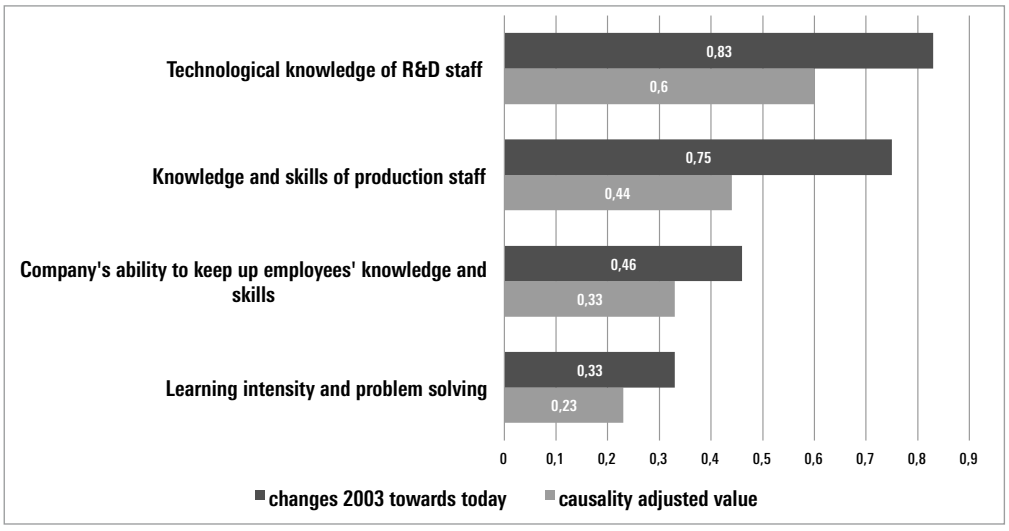


FIGURE 6 - Changes of central characteristics of innovative capacity – Human capital
LuFo III UN, n=90; differences raw and weighted

Applicability within the survey showed that on the whole the definition of TRL seemed to give a good representation of concrete technological milestones in specific projects to monitor project's progress. In single cases, though even using a general definition of TRL, the metric was not applicable to the specific project. Comments mentioning this are sporadically found in the respective fields of the survey. The reliability is also affected by the fact that it cannot be established whether the TRL assessment has been done according to a standard process used within a company, which normally is developed on the basis of the general definition, or if the project leader gave his assessment while filling in the questionnaire, relying on the definitions provided with the questionnaire. Former TRL assessments can be thought to be more reliable than the second type. Since TRL are widely used within aviation industry, it can be assumed though that using TRL to report technological progress can be seen as an overall reliable and valid indicator.

MEASURING INNOVATIVE CAPACITY

Changes in central items of innovative capacity of the participating companies (LuFo III) can be seen in the following figures. Participants gave two answers for each item. Firstly they were asked to give a judgement on the situation in 2003 and then for the situation at the time of the survey (2012). Judgements could be given on a six-point scale on each item (e.g. technological knowledge of R&D staff: „R&D staff have technological knowledge that allows research and development at an internationally outstanding level.") ranging from "strongly disagree" to "fully agree". For those questions that showed a difference in answers participants were additionally asked to assess to which extent the changes could be attributed to programme participation (six-point scale ranging from "no influence" to "strong influence").

In the figures the actual differences are shown as well as 'causality-adjusted' values. To calculate ‚causality-adjusted' values, raw differences were weighted with factors between 0 (no influence of programme participation) and 1 (strong influence of programme participation).

The results show noteworthy effects on relational capital (relations with suppliers, customers and research institutions) as well as – even though slightly less strong – on human capital (knowledge and skills of staff in R&D and production) and structural capital (organisation of R&D, cooperation between R&D and production). All these effect are attributed to the programme in significant shares. Since these variables can be described as ‚sluggish' – influencing them is complex and time-consuming – the observable effects are noteworthy.

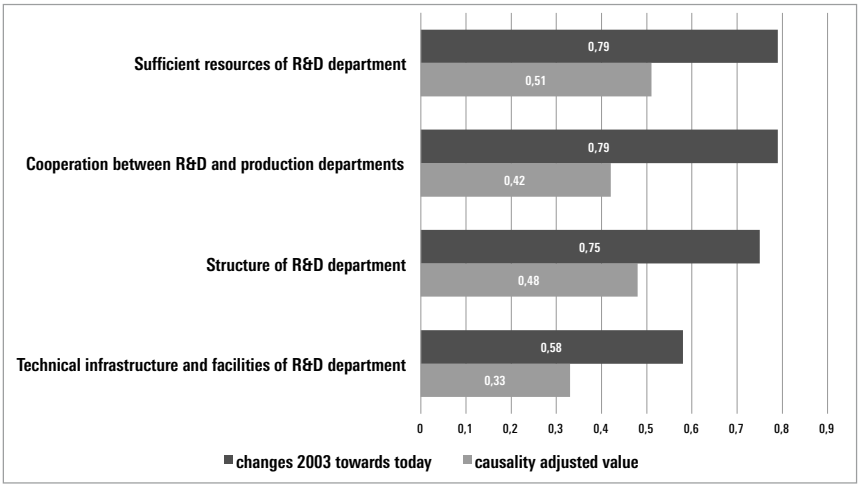


FIGURE 7 - Changes of central characteristics of innovative capacity – Structural capital
LuFo III UN, n=90; differences raw and weighted

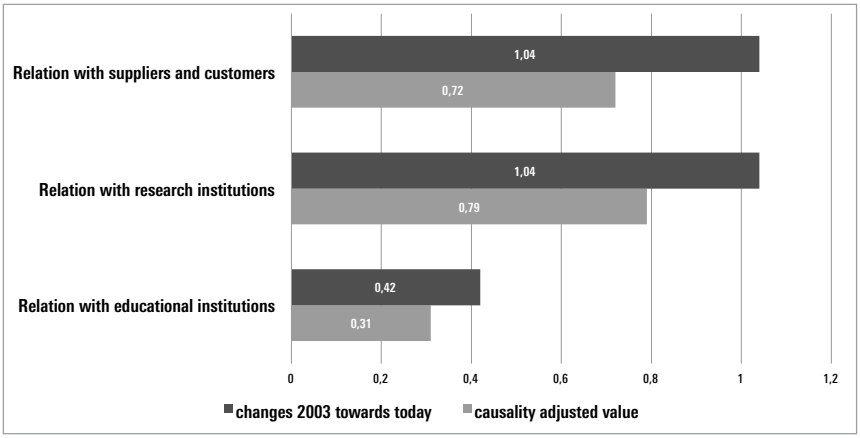


FIGURE 8 - Changes of central characteristics of innovative capacity – Relational capital
LuFo III UN, n=90; differences raw and weighted

CONCLUSION

Using TRL to measure product innovation showed that the concept was understood by participants. It could be used to show technology development and to assess single project's results. Progress of technologies can be shown at a stage where an assessment of exploitation of project results is not yet feasible.

The findings for innovative capacity are that effects could be shown which can be attributed to the programme. Considering that innovative capacity is rather „sluggish", these are surprisingly strong. Also effects on structural capital are higher than expected.

Both concepts proved to be useful within the programme evaluation. They can be applied in evaluations of programmes with very long technology development times or in evaluations that are carried out concurrently to the programme, or too early to find mentionable monetary impacts. Questions for further investigation remain, for example how to capture projects that do not fit into TRL definitions. Also the instrument for innovative capacity needs to be developed further to establish the link between innovative capacity and innovation performance in more detail.

We will use both concepts again in different areas. For TRL it will be interesting to see whether their assumed applicability to other technology fields will proof to be true.

Within Institute for Innovation and Technology an indicator system on innovative capacity has been established – based on publicly available statistical data – that additionally allows for regional, industry, and country comparisons. This concept gives a framework for comparing and benchmarking the results of a single programme.

REFERENCES

Alwert, Kay (2005): Wissensbilanzen für mittelständische Organisationen – Entwicklung und prototypische Anwendung einer geeigneten Implementierungsmethode. Produktionstechnisches Zentrum Berlin (PTZ). (Dissertation an der Technischen Universität Berlin).

ASD (R&E) (2011): Technology Readiness Assessment (TRA) Guidance. Prepared by the Assistant Secretary of Defense for Research and Engineering (ASD(R&E)). Revision posted 13 May 2011. Department of Defence. http://www.acq.osd.mil/ddre/publications/docs/TRA2011.pdf (Access date: 15.05.2012)

Bilbro, James/ Dennehy, Cornelius/ Desai, Prasun/ Kramer, Corinne/ Nolte, William/ Widman, Richard/ Weinstein, Richard (2010): Status of the development of an International Standards Organization (ISO) definition of the Technology Readiness Levels (TRL) and their criteria of assessment. 13th ANNUAL SYSTEMS ENGINEERING CONFERENCE. San Diego, October 2010. http://www.dtic.mil/ndia/2010systemengr/ThursdayTrack7_11070Nolte.pdf (Access date: 12.12.2012)

Bilbro, James/ Dennehy, Cornelius/ Desai, Prasun/ Holzer, Jenny/ Kramer, Corinne/ Nolte, William/ Widman, Richard/ Weinstein, Richard (2011): ISO TRL Definitions and Assessment Criteria – Status. 14th ANNUAL SYSTEMS ENGINEERING CONFERENCE. San Diego, October 2011. http://www.dtic.mil/ndia/2011system/13053_BilbroWednesday.pdf (Access date: 12.12.2012)

Cedefop (2012): Learning and Innovation in Enterprises. Luxembourg: Publications Office of the European Union

Cohen, Wesley M./ Levinthal, Daniel A. (1990): Absorptive Capacity: A New Perspective on Learning and Innovation, Administrative Science Quarterly, 35 (1), 128–152

DIN NL/ AST / DLR (2008): Abschlussbericht zum INS 224 Projekt: Risiko-kontrollierte Anwendung von Innovation & technologischem Fortschritt – Standardisierte Entscheidungshilfe zur Reifegradbewertung im Produktlebenszyklus – Machbarkeitsstudie. Normenausschuss Luft und Raumfahrt des DIN e.V., Astrium Space Transportation und Deutsches Zentrum für Luft- und Raumfahrt e.V. http://www.nl.din.de/sixcms_upload/media/2566/INS_224_Endbericht.pdf (Access date: 15.05.2012)

DRD/DDR&E (2009): Technology Readiness Assessment (TRA) Deskbook. Prepared by the Director, Research Directorate (DRD), Office of the Director, Defense Research and Engineering (DDR&E). July 2009. Department of Defence. http://www.skatelescope.org/public/2011-11-18_WBS-SOW_Development_Reference_Documents/DoD_TRA_July_2009_Read_Version.pdf (Access date: 14.12.2012)

Graettinger, Caroline P./ Garcia, Suzanne/ Siviy, Jeannine/ Schenk, Robert J. / Van Syckle, Peter J. (2002): Using the Technology Readiness Levels Scale to Support Technology Management in the DoD's ATD/STO Environments. http://www.sei.cmu.edu/reports/02sr027.pdf (Access date: 11.12.2012)

Jaedicke, Wolfgang/ Karl, Helmut/ Mitze, Timo/ Schwab, Oliver/ Schwarze, Kristin/ Strotebeck, Falk/ Toepel, Kathleen/ Untiedt, Gerhard (2009): Entwicklung von Performanzindikatoren als Grundlage für die Evaluierung von Förderprogrammen in den finanzpolitisch relevanten Feldern. http://www.ifsberlin.de/dokumente/file/I31%20EB5%20Dezember2009.pdf (Access date: 26.06.2012)

Luoma, Päivi/Raivio, Tuomas/Tommila, , Paula Lunabba, Johan/ Halme, Kimmo/ Viljamaa, Kimmo and Lahtinen, Henri (2011): Better results, more value. A framework for analysing the societal impact of Research and Innovation, Tekes Review 288/2011, http://www.tekes.fi/u/Better_results_more_value.pdf (Access date: 03.10.2012)

Oslo Manual (2005): Oslo Manual: Guidelines for Collecting and Interpreting Innovation Data, 3rd Edition, http://www.oecd.org/innovation/inno/oslomanualguidelinesforcollectingandinterpretinginnovationdata3rdedition.htm (Access date: 26.03.2013)

Peters, Bettina / Hud, Martin / Köhler, Christian / Licht, Georg (2012): Ökonomische Bewertung von staatlichen Investitionen in Forschung und Innovation, Studien zum deutschen Innovationssystem Nr. 15-2012, ZEW, Februar 2012

Tiwari, Rajnish (2008): Defining Innovation, Compiled by: Rajnish Tiwari, February 2008. http://www.global-innovation.net/innovation/Innovation_Definitions.pdf (Access date: 09.04.2013)

AUTHORS

CHRISTIANE KERLEN
Dr Kerlen Evaluation
E: info@kerlen.co.uk

ERNST A. HARTMANN
Institute for Innovation and Technology
E: hartmann@iit-berlin.de

EXPECTATIONS ON THE LONG-TERM IMPACT OF RESEARCH FELLOWSHIPS FROM AN EVALUATION PERSPECTIVE: CHALLENGES AND LIMITS TO MEASURE SIDE-EFFECTS

Christina Schuh

INTRODUCTION

How is the impact of fellowships defined from the perspective of policy makers, the Humboldt Foundation and evaluators? Are there differences and how can they be overcome?

These questions will be discussed by presenting the design, used methods and the relevant results of two already completed evaluation studies.

By sponsoring international research collaborations the Humboldt Research Fellowship Programme aims to gain the "best minds" for Germany as a location for research and to build and develop an enduring, world-spanning network of elites.

Beyond indicators looking for the career development of the alumni the challenge to capture careers outside academia is discussed. Additionally, research projects realized by the Georg Forster Fellowships must address issues of significant relevance to the further development of the developing and threshold countries of origin of the applicants. Methodological limits to measure this "side effect" are focused.

THE HUMBOLDT FOUNDATION'S PROGRAMME EVALUATION

The Alexander von Humboldt Foundation promotes academic cooperation between excellent scientists and scholars from abroad and from Germany. Their research fellowships and research awards allow scientists to come to Germany to work on a research project chosen by themselves together with a host and collaborative partner. As an intermediary organisation for German foreign cultural and educational policy the Humboldt Foundation promotes international cultural dialogue and academic exchange also with developing countries, emerging economies and transition states. The Humboldt Foundation places great emphasis on the evaluation, quality control and reporting of its activities. Since 2006, an independent Academic Council has steered the evaluation of the Humboldt Foundation's sponsorship programmes. The Council is responsible for monitoring and mentoring the evaluation, developing ideas based on the results of the evaluation and formulating concrete recommendations. An internal project team consisting of the programme manager and the evaluation unit proposes an evaluation design (programme goals, indicators, and methods). The project realization is generally done by external evaluation agencies. Finally, the board of trustees, including members from the funding ministries, decides on the action plan as the consequence of the evaluation.

FELLOWSHIP PROGRAMMES AND THEIR EVALUATION

Humboldt Research Fellowships offer scientists and scholars of all nationalities and disciplines the opportunity to conduct a research project of their own choice in cooperation with an academic host at a research institution in Germany. By sponsoring international research collaborations the Humboldt Foundation aims to gain the "best minds" for Germany as a location for research, and to build and develop an enduring, world-spanning network of elites.

Beyond indicators looking for the career development of the alumni, the challenge for programme evaluation is to capture career development outside academia and the side effects on cultural policy. More concretely, for example the operationalization of the image of Germany which the fellows took back to their current career step should be measured.

The results of the evaluation, done by Technopolis Austria (Katharina Warta and Anton Geyer) which employed various methods including a representative survey of research fellows and their hosts in Germany, interviews with stakeholders and a comprehensive analyses of the foundation's database endorse the mission of the Research Fellowship Programme and confirm that it achieves its objectives: Research stays are academically productive and provide a solid basis for scientific qualification and further cooperation. The majority of hosts and fellows maintain the academic contacts they have made during their stay in Germany; collaborations often go beyond the boundaries of the host institute and include other Humboldtians, too.

Today, many alumni hold leading positions all over the world, especially in academia and research, but also in politics, culture and business. On the basis of sustaining contacts and sponsoring its alumni the Humboldt Foundation fosters a worldwide network that is accessed by Ger-

man science and research institutions and used for international science and cultural policy purposes.

Scientists and scholars of all disciplines from developing countries, emerging economies and transition states (excluding People's Republic of China and India) can apply for a Georg Forster Research Fellowship. The main difference to the Humboldt Research Fellowships is, that the research proposal must address issues of significant relevance to the further development of the country of origin of the applicant and, in this context, promise to facilitate the transfer of knowledge and methods to developing and threshold countries.

The evaluation study by Arnold Bergstraesser Institute Freiburg faces the challenge of how to measure the benefit for developing countries by using a mixed-methods approach: particularly an online survey, three in depth country studies and interviews with main stakeholders of the selection committee and policy makers.

DEFINITION AND MEASUREMENT OF SIDE EFFECTS

Both evaluation studies have in common that they needed to handle the definition, operationalization and measurement not only of the programmes' impact on research and academic careers but also on so called side effects.

The Humboldt Research Fellowship Programme addresses the intellectual elite worldwide from all scientific disciplines and aims to support important careers inside and outside academia. The Georg Forster Research Fellowship Programme addresses careers with significant relevance to the further development of the country of origin of the alumni. Whereas the main effect in both programmes promoting research stays in Germany can be seen as the scientific impact on the academic career of the fellow the side effects differ: for Humboldt it is the expected important career outside academia in politics, culture and business and for Georg Forster it is a career with a significant relevance to the further development of the country of origin.

Besides the operationalization of these side effects for the programme's evaluation it is important to discuss the relation between these main effects and side effects: are the main effects necessary and sufficient for the programme's success or does the programme's success also depend on the so called side effects?

The Humboldt Foundation's selection process can give insight on these relations: for Humboldt, academic excellence is the sole selection criterion that means that the main effect should be necessary and sufficient; for Georg Forster the research project must have an influence of significant relevance for the development of the country besides academic excellence, which means, that both the main effect and the side effect are necessary for the programme's success.

RESULTS

For the Humboldt Research Fellowship the two evaluated programme objectives facing the so called side effects are[1]:

- Programme Objective 4 (PO4): To facilitate access to international experts and decision-makers in the fields of academia, politics, culture and business for partners from the relevant sectors in Germany.
- Programme Objective 5 (PO5): To convey a multifaceted, realistic image of Germany by creating personal and cultural bonds, breaking down prejudices and promoting knowledge of the science system.

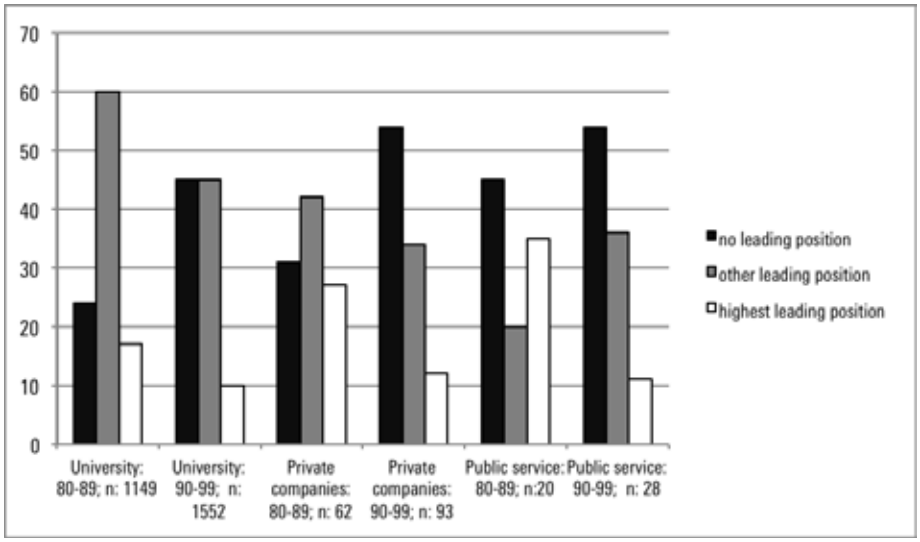


FIGURE 1 - Humboldt-Alumni in leading positions, regarding main fields, accepted cohorts 1980-1999 in %

1 The complete evaluation report (in German) can be downloaded from the Foundation's website: http://www.humboldt-foundation.de/pls/web/docs/F25500/ergebnisse_evaluation_hfst_lang.pdf

One indicator for PO4 was the question, whether the alumni reach leading positions or not.

Results in figure 1 show that the majority of the alumni stays at university and can reach leading positions (additional 578 alumni stay in research but not at universities and 48 work as self-employed).

Some examples for the operationalization of the multifaceted, realistic image of Germany described in PO5 are shown in figure 2.

Although, not every alumni answered each question, there are some concerns about the image of Germany the researchers took back to their next career step, like for example "dual career opportunities" (50% judged dual career opportunities as surprisingly negative) or the "experience of hostility" (about 20% got in contact with hostility). Childcare, infrastructure/ public transport and the knowledge of English of the population were seen surprisingly positive.

For Georg Forster the so called side effect is expressed in Programme Goal 3[2]:

Support of qualified researchers from developing and threshold countries in their role as important change agents in reform processes in research, economy, politics and society.

Some relevant Indicators investigated in the evaluation are:

- self-assessment of the relevance of the own research and publications for the development of the country
- counselling and co-working with important institutions in research, economy, politics and society
- reviewer and counsellor for development cooperation

Table 1 shows results from an online-survey of the alumni considering the influence of the Georg Forster Fellowship Programme on Fellows and their country of origin.

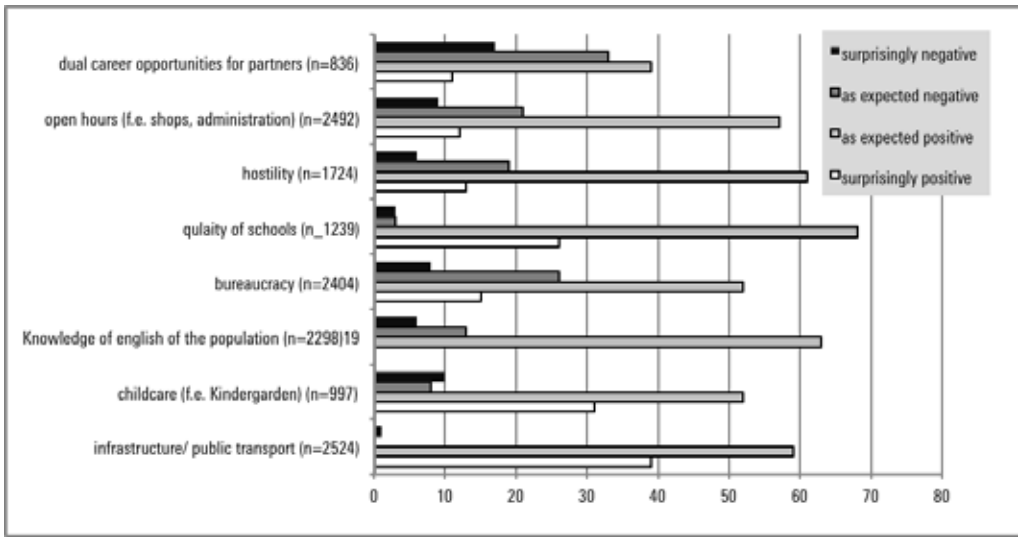


FIGURE 2 - Experiences in Germany (Excerpt) in %

Possible effects	Very important	Important	Less important	Not important	N
Improving my academic knowledge	79%	20%	1%	*	298
Improving my methodological skills	72%	24%	4%	*	295
Improving my capacity to act as a multiplier in Higher Education and Research	68%	28%	4%	*	299
Understanding a different culture of knowledge (management)	51%	37%	11%	1%	297
My research results contribute to solving development challenges	45%	42%	11%	3%	295
Positive impact on my activities in other fields (civil society, social/cultural/political activity, etc.)	42%	36%	19%	3%	298
Relevance for crucial economic, social and political problems in my country	34%	35%	22%	9%	297

TABLE 1 - Influence of the Georg Forster Fellowship Programme on Fellows and their country of origin, n = 299

2 The evaluation report will soon be available on the Humboldt Foundation's website: http://www.humboldt-foundation.de/web/evaluation.html

The most positive effects of the fellowship were seen in the improvement of the academic knowledge and in methodological skills. Also (very) important is the improvement of the capacity to act as a multiplier in Higher Education and Research. Still more than two thirds of the alumni see it as an (at least) important impact of the fellowship that their research results contribute to solving development challenges or have a positive impact of their activities in other fields (civil society/social/cultural/political activity, etc.) or relevance for crucial economic, social and political problems in their country.

Although the main impact can be seen on academic benefits, the so called side-effect was assessed at least important by more than two thirds of the alumni.

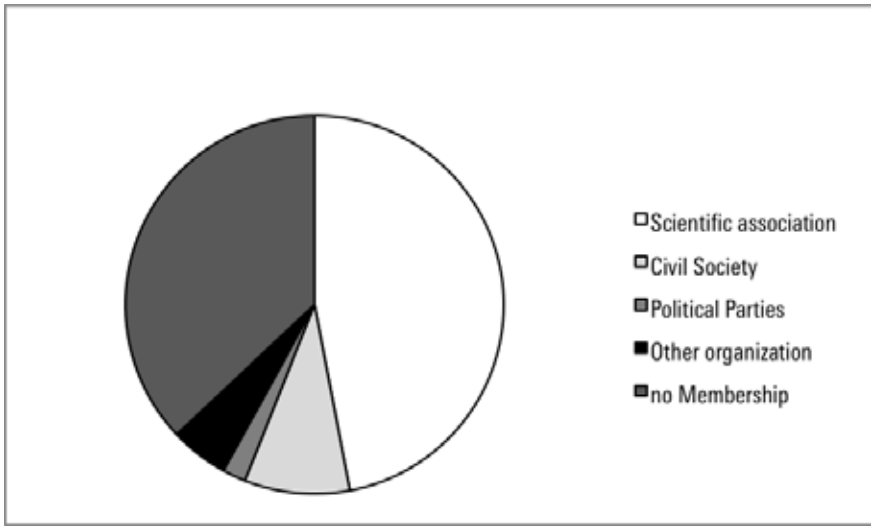


FIGURE 3 - Membership in different associations in %; n=299

Another relevant indicator for side effects in the Georg Forster Fellowship Programme is the membership in different associations shown in figure 3.

As expected, most of the alumni (50%) are organized in scientific associations. But about 16 % are members of associations in the field of civil society, political parties or other subjects, so that a special interest in issues others than science can be assumed. But of course, there is a lack of data to compare these results with the worldwide scientific community.

LESSONS LEARNT

The evaluation of the Fellowship Programmes made a common understanding of the definition and relation of main effects and side effects for the different perspectives (of the funding ministries, evaluators and the Humboldt Foundation) necessary.

Besides a discussion of the programme goals there must be a dialogue not only about relevant indicators to measure them but also about the importance of the different goals and their relation. While the definition of the programme goals and the belonging discussion between funding ministries and Humboldt Foundation is generally done when the programme is invented, the discussion about indicators and the importance of goals and their between-relation is done during the evaluation process. The selection process and the weighting of different criteria during the decision whether an applicant receives the fellowship or not, can give useful insight for the weighting of the different programme goals in the programme evaluation.

Also a dialogue about reasonable indicators to measure goals, main and side effects is important. Since benchmarks are rare a common understanding between the different parties of what is a lot and what is not enough to reach a goal is necessary for each indicator.

Finally, as career development is very heterogeneous the aggregation of data risks to lose a lot of information which made a mixed method approach with case studies included for a good data interpretation indispensable.

AUTHOR

CHRISTINA SCHUH
Humboldt Foundation
E: Christina.Schuh@avh.de

CONFERENCE "EVALUATION OF STI POLICIES, INSTRUMENTS AND ORGANISATIONS: NEW HORIZONS AND NEW CHALLENGES", **SESSION SUMMARIES**

To enable students from the field of science/technology/society-studies access to the conference "**Evaluation of STI policies, instruments and organisations: New horizons and new challenges**", fteval cooperated with the Department of Science and Technology Studies (S&TS) of the University of Vienna. fteval granted free conference permission for 15 students, but they were obliged to document the major topics and take-away-messages of each session in return. In the following section you will find their impressions, reflections and perceptions of several conference sessions. We included also two more comprehensive panel summaries drafted by Mario Steyer (bmvit – Austrian Federal Ministry of Transport, Innovation and Technology) and by Martina Lindorfer (ZSI – Centre for Social Innovation).

With these few session summaries we hope that you can catch a glimpse of what happened during the two intensive conference days in Vienna, packed with inspiring discussions and enlightening presentations!

TABLE OF CONTENTS

FIRST KEY NOTE SPEECH

WALTER AIGNER

The first key note speech by Dr. Wilhelm KRULL, Volkswagenstiftung had the title "The Politics and Policies of Evaluation in a Multi-level Research System". It did not follow an explicitly mentioned theme but provided reflective remarks around a cover story – earlier this year – published in the Economist "How science goes wrong" (October 19th 2013) as well as on the ongoing activity "How to establish a European research area" .

KEY CHALLENGES

Challenges raised (in chronological order from the presentation; no emphasis or hierarchy was given in the keynote):

- Ambitiousness versus over-ambitiousness
- Evaluation versus over-evaluation
- Some unintended consequences from evaluation methodology
- You learn more from failure [elliptical statement probably claiming "but hardly anyone dares to risk failure"]
- Variation in evaluation criteria across disciplines (examples given: humanities focus on originality; social science on new methodology)
- Reviewers overestimate their openness toward new issues or approaches. (Some of the most original people never hand in proposals.) Proposers are discouraged by statements like "we never had such a proposal before".
- Predominant approach of two-day evaluation meetings leads to nothing else than nice peer-to-peer networking.
- True evaluation can only be done after 8 years – however policy cycles and project duration request researchers and public administration to come up with follow up proposals before they can possibly know intermediate results from their previous activities or research projects.
- Transparency paradox: the more publicly available documentation and transparent evaluation processes a funding body has implemented the higher is the inclination to hide failure, mistakes, mismanagement or severe problems.
- Need for experts from academia – however, during the last years, only university presidents became ministers or evaluators of strategic STI initiatives.
- Mismanagement: inappropriate deterministic concepts of evaluators
- Mismanagement: European practice of focusing on fair regional distribution of funds instead of innovation or excellence.
- Economist cover story: How science goes wrong: (50 per cent of results presented in life science research papers cannot be reproduced).

KEY ARGUMENTS

Model of 7 Cs:

1. Competence
2. Creativity
3. Commitment
4. Communication
5. Co-operation (e.g. Max Plank Institutes with local universities)
6. Continuity
7. Centers / Clusters

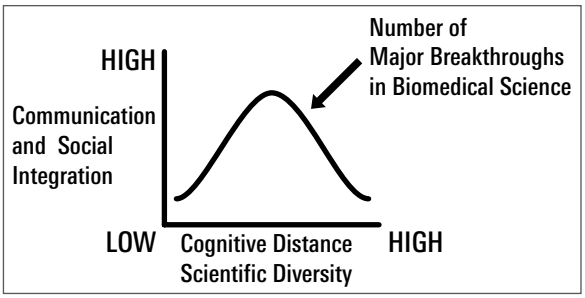
KEY RECOMMENDATIONS

Two implicit recommendations can be derived from the prominent mention of two practices within the Volkswagenstiftung:

1. VW-Stiftung pyramid approach in evaluation
2. VW-Stiftung has always a gate for those off the beaten track.

And most probably to follow the 7 Cs model?

Possibly also to adhere to the logic in Rogers Hollingsworth's two dimensional concept chart



Degree of communication vs. cognitive distance / diversity:

OPEN QUESTIONS AND DISCUSSION

There was no discussion after this key note speech. However the presentation ended with the following quotation from Samuel Beckett (Worstward Ho (1983):

"All of old. Nothing else ever. Ever tried. Ever failed. No matter. Try again. Fail again. Fail better."

PANEL 1: NEW RTI INSTRUMENTS – NEW INTELLIGENCE REQUIREMENTS: WHAT HAS TO BE CONSIDERED IN RTI EVALUATIONS?

MARIO STEYER

Panel: Katharina Warta, Technopolis (Moderator); Dominique Guellec, OECD; Stefan Kuhlmann, University of Twente; Rupert Pichler, Austrian Federal Ministry of Transport, Innovation, and Technology

KEY CHALLENGES

In a world of growing complexity, the complexity of recognition patterns of systems is also - nolens volens – growing. The field of research, technology and innovation is not different. There are numerous ways to deal with economic and societal challenges concerning RTI, principles and methods of funding. Accordingly, there are numerous ways to measure the success and consequences of actions taken in the field of RTI, i.e. instruments of evaluation.

So what instruments should we 'pick' to evaluate the given measure? Are some instruments simply (and always) better than others, or is it all about context? Quoting the subtitle of Panel 1:

What has to be considered in RTI evaluations?

Following Stefan Kuhlmann, 5 points are of importance if we want to take the big picture.

- First, we need to reach a point beyond the evaluation hype. Today we have a far better understanding of processes and evaluation itself but at the same time we have come to a point where quantitative methods stand for the 'gold standard' of evaluation (even though often not properly used) and evaluators and policy makers have to become more self-critical. Evaluation is not perfect just because it's done.
- Second, complexity is and remains high, the interaction of interdependent participants in the field increases steadily.
- A third point considers the difficulty of the evaluation of grand challenges, which is to some extent related to the second point.
- The fourth aspect concerns the underestimating – that is the meaning and role – of new actors. Societal groups affected by policy recommendations need to be taken into account.
- Finally, there are new dimensions for evaluation, such as responsible research and the performance of RTI in a global perspective.

KEY RECOMMENDATIONS

From the policy maker's perspective, Rupert Pichler stresses the need for evaluation and strategic intelligence as a means to cope with complexity. RTI is a particularly difficult policy field as it almost entirely relies on the input of experts rather than on the feedback of a wider group of citizens as clients or customers of public services. Therefore, Pichler challenges Kuhlmann's claim for modesty: now that politicians have eventually come to understand the importance of evaluation it would be fatal if evaluators themselves renounced the expectations they had raised in the past. Accordingly, evaluation must keep its ability to communicate with the policy system and should not retreat into the safe haven of academia as if it were a genuine scientific discipline.

Can a pragmatic access to all these challenges be of use? An example would be the evaluation of R&D tax credits, actually very popular in Europe and OECD countries – a rather simple instrument but no comparable systems of funding can be found. One solution could be: Use scientific disciplines that are needed to face evaluation problems, use evaluation in a "giving advice"-manner and don't tell the program owners or policy makers how to implement all given advices. Dominique Guellec is convinced that these guidelines and the appropriate choice of instruments for evaluation (regarding size, dimension and general conditions/restraints of the funding program/scheme/regime) will help to deal with specific national aspects of funding and the information flood that evaluators can be confronted with.

OPEN QUESTIONS AND DISCUSSION

The result of the first panel was obvious: The emerging requirements for 'good' or even 'perfect' evaluation cannot be defined, but not all hope is lost. The need for a new modesty was claimed, but this could be dangerous: it could destroy the culture of trust, which for example in Austria has been established over the last 15 years of evaluation work – done together.

Failure of programs or initiatives has to be accepted, every side should learn to distinguish between strategic and 'regular' evaluations. To know the limits of each evaluation carried out helps to prevent misunderstandings before the evaluation has even started. Finally, clear and simple statements are needed, for both sides.

LUNCHBREAK PANEL: ENHANCING THE USE OF RESEARCH INFORMATION SYSTEMS FOR RTI EVALUATION

MARTINA LINDORFER

Representatives of agencies that have in-house research information systems discussed about challenges and opportunities of their systems and tools for the monitoring and evaluation of national research activities, outputs and impacts. Three dimensions, that are essential for making huge data available and useable, were put up for discussion: the technical dimension (How does a good research information system look like?), the nature of data (What kind of data shall be collected?), and the method of data collection.

Panel: Sybille Hinze, ifq – Institute for Research Information and QA (Moderator); Gretchen Jordan, Innovation LLC; Göran Marklung, Vinnova; Sabine Mayer, Austrian Research Promotion Agency; Christina Schuh, Humboldt Foundation

KEY CHALLENGES

The key challenge, that became most evident during the discussion, was how to find the right balance between enhancing comparability of data (for customization) while at the same time keeping up specificity of data, which is crucial to an agency. This entails other general questions, such as the most appropriate classification of data, which is still an unsolved problem.

Many different approaches of data management are being applied across countries and agencies (some of which were presented in the session). For the sake of better alignment and comparability of data in the future, good practices shall be identified (see http://www.eurocris.org/Index.php?page=homepage&t=1) and be followed on a more global scale. However, good practices are often designed according to very specific needs of funding agencies and the demands of their clients, which are difficult to handle on a supra-institutional scale. Another key challenge, which came up in the discussion, is the fact that research information systems are still very focused on the "input side" and that good methods for the inclusion of research outcomes, effects and impacts do not yet exist.

Still a lot of conflict has to be passed in the process of aggregating data and finding common indicators. Because the categorization of information entails a certain simplification of information and the partial loss of specificity of data, it is an emotional and uneasy task to come to compromises.

KEY DATA

Sabine **MAYER** (Austrian Research Promotion Agency) presented the specificities of the Austrian Research Promotion Agency's information system, which is one database for all national funding implemented by the Research Promotion Agency. The system is constantly being improved, but several key questions are not solved yet, such as: What is the most appropriate categorization of data? What is a good method to include project outputs, outcomes and qualitative information? How to ensure access to data for an interested public when a considerable part of information is confidential?

Christina **SCHUH** (Humboldt Foundation) reported on the new research information system that shall be implemented by the Humboldt Foundation next year. More than providing good funding statistics, the system wants to make data comparable on an international level. The problem here is that data coverage is not given for many countries. Ms. Schuh also referred to the difficulty of including career data and keeping it current and complete; as well the question of how to categorize academic prizes is not solved. A more general open question is which indicators shall be monitored on a continuous basis.

Gretchen **JORDAN** (Innovation LLC) commented on the STAR METRICS project, a partnership, which was initiated five years ago in the US between science agencies and research institutions to document the outcomes of investments in science. The importance of STAR METRICS lies in enhancing comparability of data across US funding agencies, as the US research system is much decentralized. STAR METRICS is a two level system, which collects information on research jobs on level 1, and links input measures to research outcomes on level 2. This second level is seen very critical by Gretchen Jordan, because according to her, neither the data on research outcomes, nor a strategy for linking inputs and outcomes are available at that point.

KEY ARGUMENTS

The discussion was mainly driven by open questions and unsolved challenges, rather than by clear statements.

To name only some arguments that were raised:

- Organizational change is necessary to increase the comparability of data (across projects, issues, etc).
- The quality of the standing data is crucial.
- International benchmarks need to be established for better comparability.

KEY RECOMMENDATIONS

Again, it was foremost stressed, that for many methodological and structural problems solutions are not yet found, and that the agencies are still struggling with making the right choices. Some recommendations were however made:

- Keep the research information system simple
- As an agency, don't make own definitions
- Find a good way to keep data updated & clear
- Assure a very stable base for the analysis of data
- Avoid losing data by using too tight structures

OPEN QUESTIONS AND DISCUSSION

Main open questions were already summarized in the "key challenge" section. Key points/ questions of discussion may be summarized as:

- How to make data useable for RTI evaluation?
- How to make data comparable across different issues, needs & institutional settings?
- Is there an optimal way of collecting, categorizing and updating data?
- How to keep the balance between customization and being faithful to details?
- How to include qualitative information and impact indicators?

SPECIAL SESSION: HORIZON 2020 – THE OVERALL EVALUATION APPROACH: CRITICAL REFLECTION AND DISCUSSION

TUDOR B. IONESCU

Impulse presentation by Peter FISCH, European Commission

Panel: Jakob EDLER, University of Manchester - Manchester Institute of Innovation Research (moderator), Peter Van den BESSELAAR, VU University Amsterdam, Erik ARNOLD, Technopolis

SUMMARY

AGENDA

What is new / Challenges / Problems / Discussion

Presentation of Dr. Fisch

EU FRAMEWORK PROGRAMMES CHARACTERISTICS'

- Figures:
  - Over 100k applications / 20k funded
  - 39% University, 29% Industry, rest Research institutions
- Annual monitoring reports are a novelty in FP7
- Statistics about where researchers publish have improved.

HORIZON 2020 BASICS

- Unifying three different programs (FP, CIP, EIT)
- Coupling research to innovation
- Three pillars
- Innovation indicators – in the legal text, list of key indicators for assessing achievement of Horizon 2020
- Achievement indicators:
  - Academia: number of publications, ranking of publications (Journals, conferences)
  - Industry: patent applications…
- Annual monitoring of cross-cutting issues
- Challenges
  - Data – how to get data in databases and meaning out of data?
  - Store data in open access area
  - Link databases with Scopus, Web of Science
  - How to integrate these data with social networks and other databases?
  - Technical feasibility
  - Political agreement
  - Legal issues (confidentiality, personal data…)
- Methods of analysis:
  - Counterfactual analysis
  - Formal consultation of stakeholders
  - Survey techniques using social media
- Improve the communication process

DISCUSSION ROUND

Eric Arnold

- Proud about progress in evaluation of FP programmes
- What do patents really mean? – we need to look for a sensible use of that
- Trivialisation of bibliometrics
- "Fascination" about open access:
  - Public sector will not publish if it costs
- Evaluation is not a science. It is a way to use robust methods as good as we can.

Peter Van den BESSELAAR

Q: WHY DID FISCH USE THAT PARTICULAR SET OF INDICATORS THAT ARE VERY GENERAL / THEY WORK FOR ANY PROGRAMME / THEY DON'T REFLECT ANY SPECIFICITY OF THE FRAMEWORK PROGRAMMES… WE NEED A LOT OF THINKING ABOUT HOW TO EVALUATE DATA.

A (P. Peter):

- Key performance indicators (KPI) are the results of a painful process. KPI must reflect standard activities. Anything else was rejected because too specific à unable to compare with other things.
- Interesting data in terms of network patterns of collaboration à increased focus on this rather than patent data in Horizon 2020

Q1: WHAT ABOUT TEXTUAL DATA? WILL THEY BE OPENED? CAN THEY BE USED FOR EVALUATION?

A (Peter): Yes and no…. more reports for consortia on management and publications. Producing final reports is a waste of time. Who reads them? – They will analyse them with software like that for detecting plagiarism. Big data is a challenge.

Q2: RESPONSIBLE RESEARCH AND INNOVATION? WHAT IS THAT?

A (Fisch): What is the opposite of that? Irresponsible research? The substance of this is to look at the interaction of science with society.

SESSION 1: NEW APPROACHES FOR EVALUATING STI POLICIES AND INSTRUMENTS

EMIL ANGELOV

The main topic Session 1 was presenting various researches attempting to provide frameworks or guidelines for evaluation and measuring of STI policies' impact and its instruments.

PRESENTATIONS

Matthias WEBER, AIT and Wolfgang POLT, Joanneum Research: *Assessing mission-orientated R&D programs: combining foresight and evaluation*

Pierre-Benoit JOLY, Institut Francilien Recherche Innovation Société (IFRIS) and INRA/SenS, Ariane GAUNAND, Grenoble Applied Economics Lab (GAEL), Philippe LAREDO, IFRIS, Mireille MATT, GAEL and Stéphane LEMAIRE, IFRIS: *Designing and implementing a new approach for the ex-post assessment of impact of research – a return of experience from the ASIRPA project*

Stephanie DAIMER, Fraunhofer Institute for Systems and Innovation Research: *Evaluating the novel German "VIP" measure – addressing the stage of translational research between basic research and valorisation*

Stephan ROPER, University of Warwick/ Warwick Business School: *An experimental approach to industrial policy evaluation: The case of Creative Credits*

Abdullah GÖK, University of Manchester/ Manchester Institute of Innovation Research: *The Use of Experimental and Quasi-Experimental Methods in Innovation Policy Evaluation*

KEY CHALLENGES

Probably the main challenge of STI policy evaluation lies in its all-inclusiveness and transformative intention targeting social, economic, technological and environmental objectives which are further made difficult by complex management requirements and network interactions. The new mission orientation has developed from the classical model to a new frame for designing policies that we have today – with wider impact, inclusiveness of social innovations and socio-technical solutions and a multidisciplinary perspective that requires international collaboration. Furthermore, there is a growing requirement for science to provide quality evidence on the impact of policy interventions and a tendency to see evaluation as an instrument that would bring closer and provide better understanding of all stakeholders in the STI arena – policy makers, clients and evaluators. In order to meet these demands new methodological approaches are being developed.

KEY DATA

All presentations, apart from one, were backed-up by research data obtained from different, sometimes experimental approaches including methodologies of standardised case studies, longitudinal data collection strategies of qualitative and quantitative data and experimental and quasi-experimental designs. In many cases the researchers were trying to break new ground in order to answer the increasing complexity of STI policy interactions and to provide multi-perspective approach and basis for learning.

KEY ARGUMENTS

STI policy assessment is a new form of accountability towards society. If done properly it will translate the potential of research funding into knowledge and technology transfer and into tangible benefits; it will induce learning in the broader stakeholder community and it will provide new rationale and more in-depth understanding of long-term impact of policy to policy makers and other stakeholders.

KEY RECOMMENDATIONS

Depending on the presentations there were some specific recommendations but the overall impression was that there are two aspects of STI policy evaluation that resonated through the session: First of all, a comprehensive evaluation policy is needed that would take into consideration all the challenges of policy impact assessment (diffusion, time-horizon, attribution, multi-policy, multi-level…) that would overcome evaluation problems in all policy areas and will add issues of scientific, social and environmental issues to economic feasibility. Second, ex-ante evaluation seems to be preferred to ex-post evaluation in STI policy because it enables analysis of the anticipated impacts, optimizes the structure of a programme, the sequence of priorities, as well as the external and internal coherence of a programme. It also provides justification for the allocation of funds after all other aspects have been taken into consideration.

OPEN QUESTIONS AND DISCUSSION

The discussion during this part of the session was mainly about the ways of combining experimental and quantitative designs to emphasize the impact of evaluation and the very narrow, yet powerful, impact of such methodologies and the question of leverage in terms of ex-ante and ex-post evaluation because policy-makers are reluctant to support ex-ante evaluation because they lose some power – while ex-post evaluations are largely used to justify policies already being implemented.

SESSION 2A: ASSESSING THE VARIETY AND LONG-TERM IMPACT OF RESEARCH

VOLKER ZIEGLER

PRESENTATIONS

Erik ARNOLD, Technopolis and University of Twente, and Terttu T. LUUKKONEN: *The Research Institute of the Finnish Economy: How to evaluate research funding organisations*

Rodrigo COSTAS and Erik Van WIJK, Centre for Science and Technology, University Leiden: *Bibliometric study of FWF Austrian Science Fund (2001-2010/11): main results*

Ralph REIMANN, Austrian Science Fund: *Bibliometric study of FWF Austrian Science Fund (2001-2010/11): from the funder´s perspective*

Chris L. S. CORYN, Western Michigan University: *Central Findings and Lessons Learned from an Evaluation of the Swiss National Science Foundation*

The first lecture put emphasis on cross-national comparison of the impacts of policies that led to certain developments within each country's scientific and educational landscape. In particular, the funding structure of scientific organisations was analysed and the specific characteristics of each country's national situation was accounted for. Scandinavian countries were compared to „big players" like the United States or China.

The following lecture was held by Rodrigo Costas and it contained a bibliometrical analysis of scientific outputs recorded in the FWF (Österreichischer Wissenschaftsfonds, which funds basic research) publication system. The researchers described the way they were approaching the matter and which methods they used. The main focus was put on the citation system. One of the key messages of the lecture was to keep in mind the peculiarities that different citation systems contained. They addressed challenges they would face in the future such as the possible ways of how to bibliometrically study funding organizations, namely via data collection or benchmark analysis of funding organizations. Both ways, according to them had pro's and con's and both methods tend to overlap when applied. The researchers further described what data were deemed relevant by them and how they approached the key questions they faced methodologically. They deduced the significant role that the FWF plays both nationally and internationally and that its impact in most fields of science was high.

The lecture of Costas was followed by a very brief lecture by Ralf Reimann, who is working for the FWF, which provided a different point of view to the one of Costas. Different factors such as project duration, funding acknowledgement and different document types were addressed. Problematic aspects like „fuzzy" acknowledgement (multiple funding sources) and different „importance" of reviews compared to articles were displayed.

The closing lecture of the 2a series was held by Chris L. S. Coryn, who represented a multidisciplinary team of predominately American scientists who were commissioned and authorized by the SNSF (Swiss National Science Foundation, the biggest science agency in Switzerland) to provide a study about the efficiency of the organization and to contribute conclusive remarks regarding quality and transparency. The research team found various inefficiencies and described – besides the fact that the organisation underwent fundamental changes in their peer review system at the time of the evaluation - also the cultural differences they faced while completing their task (e.g. the cultural variation of the dichotomy evaluator – interviewed subject; cultural divide like language problems and the political interferences the research team faced). Coryn made initial remarks of how well the project was funded and presented the approach the team took to find appropriate answers. They centered their analysis towards six focal points and a number of secondary questions.

SESSION 2B: ASSESSING THE VARIETY AND LONG- TERM RESEARCH IMPACT

DUŠAN MARČETIĆ

PRESENTATIONS

Mika NIEMINEN, Kirsi HYYTINEN, Anu TUOMINEN, Heidi AUVINEN and Juha OKSANEN, VTT Technical Research Centre of Finland: *Understanding "complexity": future oriented impact assessment in complex decision-making environment*

Federica ROSSI, Birkbeck, University of London, Annalisa CALOFFI, University of Padova: *The long term effects of policies in support of innovation networks*

Ariane GAUNAND and Mireille MATT, Grenoble Applied Economics Lab (GAEL), Stéphane LEMARIE and Amandine HOCDE, Institut Francilien Recherche Innovation Société (IFRIS), Elisabeth De TRUCKHEIM, INRA: *How does public agricultural research impact society? Towards a characterization of various patterns*

Claudia MICHEL, University of Bern, Centre for Development and Environment CDE, Simon HEARN, Overseas Development Institute ODI, Gabriela WUELSER, Swiss Federal Institute of Technology Zurich, Environmental Philosophy, Thomas BREU, University of Bern/CDE: *Assessing the impacts of transdisciplinary research in reducing poverty: the case of the NCCR North-South*

Laurence COLINET, INRA, Pierre-Benoit JOLY and Philippe LAREDO, Institut Francilien Recherche Innovation Société (IFRIS), Ariane GAUNAND, Grenoble Applied Economics Lab (GAEL): *Public research organizsations and their impact on public policy from observations towards the characterization of impact*

The general topic was to see the impact that science/research (policy) has on society, therefore it was assessing the different relations between science and policy as well.

KEY CHALLENGES

The key challenges are to observe different impacts of science and research on society.

The first presentation was about how support for innovation projects can support the realization of innovation and tried to define the best line of action for one particular region.

The second one assessed the impact of public agricultural research on society.

The third presentation assessed the diverse impacts of transdisciplinary research on poverty reduction (case North- South) and how to identify the desired behavioral changes.

The fourth and the last presentation, like the second one, dealt with the impact of public agricultural research and one of the main challenges here was to evaluate how research contributes to public policy.

KEY DATA

The presentations were based on different qualitative and quantitative data. The methods employed were also different, comprising case studies (regional government in Tuscany, Italy), performance evaluations, assessments based on database descriptions and variables, classifying with different methods (Condorcet votes), rapid outcome mapping approach etc.

KEY ARGUMENTS

The key argument, the main thread common to all these presentations, as well as the afterwards discussion, was that it is no longer enough to assess the impact of science, research and development on society in the ways we used to know it (basic results, simple input/output linear model) but instead, it is necessary to consider the wider societal, political, environmental, cultural impact. Innovation must be linked to sustainable development. An evaluation of research organizations faces challenges through various impact dimensions.

KEY RECOMMENDATIONS

The main recommendations were:

- trying to have a contract between science and society
- criticizing quantitative indicators
- promoting more active involvement of policy makers in reducing the policy constraints
- research should have an impact on society, but considering which of the many existent expectations are reasonable

SESSION 4A: CHALLENGES IN ASSESSING NEW EUROPEAN RESEARCH AREA POLICIES, PROGRAMMES AND INSTRUMENTS

MARLENE ALTENHOFER

PRESENTATIONS

Emanuela REALE, CERIS CNR Institute for research on firm and growth, Maria NEDEVA and Thomas DUNCAN, University of Manchester/ Manchester Institute of Innovation Research, Emilia PRIMERI, CERIS CNR: *Assessing the impact of joint and open research programmes: a process-centred approach*

Martin-Felix GAJDUSEK, ZSI – Centre for Social Innovation and Nikos SIDIROPOULOS, University of Athens, Centre of Financial Studies: *Monitoring and Evaluation in joint calls of "horizontal – INCO" ERA-NET and ERA-NET PLUS actions*

Martin MAREK and Erich PREM, eutema Technology Management GmbH & Co KG: *Visualizing programme participations with interactive maps*

Karel HAEGEMAN and Mathieu DOUSSINEAU, Institute for Prospective Technological Studies, Joint Research Centre, European Commission: *Bridging the innovation gap: Private sector involvement in public-to-public R&D funding co-operation*

As the session title suggests, the presenters addressed challenges, problems, effects and (possible) impacts when it comes to various policies, programmes or instruments from a European research perspective.

KEY CHALLENGES

The overall theme of the session was the various challenges that may come up with regard to new European Research Area level policies and instruments, what role impact may have, and how impact can be evaluated and interpreted. The presenters addressed possible approaches and also assessment criteria that may be of importance regarding future evaluations of R&D programmes.

Emanuela Reale, the first presenter, looked at the impact of joint and open research programmes; the main challenges were how to approach the relevance of impact in research programmes, and what problems assessing impact may bring along. These are for example methodological problems (as measuring and attributing impact can be problematic, for example because of time lags), ontological problems (as the relationship between the "impactor" and the "impacted" is not direct and there may be interfering factors), or axiological problems (e.g. the outcome may have a very vague relationship to signals and actions from the scheme).

The second speaker, Martin-Felix Gajdusek, focused on monitoring and evaluations of joint research program initiatives such as ERA-NET and ERA-NET PLUS, in order to understand the dimensions of transnational/joint funding and also to improve the feasibility of monitoring and future evaluations. Main challenges mentioned by the presenter were the often insufficient documentation on national and regional bases or different regional funding rules, which complicate matters of evaluation and monitoring.

Erich Prem, the third presenter, addressed the challenge of visualising programme participations with interactive maps (so e.g. total public funding per country, per number of project, or per type of organisation). One of the challenges regarding such tools can be the misuse of them (see also *6: Open questions and discussion*).

The fourth speaker, Karel Haegeman, addressed private sector involvement in public-to-public R&D funding, which, according to Haegeman, may help to solve societal problems (e.g. health issues, food production, agriculture and water, energy issues, ...) and improve competitiveness. In that sense, Haegeman introduced these societal problems as key challenges which may be solved through a stronger private sector integration in public-funded research.

KEY DATA

The key data used for the first presented project was data of two (out of ten) cases from the JOREP Project (a project mapping joint and open research programs funded by the European Commission). The second speaker based his presentation on data collected through a survey in different target countries. For the third presentation, data on funding from joint technology initiatives of the EU was used in order to present the features of the visualisation tool developed by the strategic research and technology consultant company eutema. The fourth presentation was based on information of publicly funded (EU) R&D, with a special focus on if/how the private sector is already involved in programmes addressing societal challenges.

KEY ARGUMENTS

The key argument made by Emanuela Reale was that looking on the impact of research programmes might improve evaluation (processes), and enhance reflexivity, which means "to scrutinise the relationships between the actual requirements of a system, activities and results, and the ultimate change the policy instrument wants to achieve".

In Martin-Felix Gajdusek's presentation, the key argument was that monitoring and evaluation processes may help to identify practical

dimensions of joint funding, which might also improve the feasibility of monitoring and future evaluations.

The interactive map tool presented by Erich Prem vividly demonstrated how data (from funding initiatives) can be shown in a dynamic and interactive way, which very much distinguishes it from traditional maps as it gives a much broader range of possibilities to represent information.

The key argument made by Karel Haegeman is that combining private sector and public-to-public R&D funding on various levels of research and development may improve addressing and solving societal challenges and may increase competitiveness.

KEY RECOMMENDATIONS

The speakers addressed different recommendations, which were already partly discussed above. In that sense, Emanuela Reale recommends that looking on the impact of (joint) research programmes may enhance evaluation of those and reflexivity. Martin-Felix Gajdusek suggests that an improved and clear documentation or concrete program objectives may improve monitoring and evaluation processes. Erich Prem demonstrated and recommends that using interactive tools for presenting data "can give you a more powerful way to give answers" and that conveniently visualised data may improve the understanding of the presented information. The key recommendation by Karel Haegeman is that the involvement of private sector funding in publicly funded research can help solving societal challenges and improve (international) competitiveness.

OPEN QUESTIONS AND DISCUSSION

In the discussion after the four presentations, a couple of issues were addressed both by the discussant, Jakob Edler, and the audience. Edler stated that the first and the third presentations (Reale and Prem) could also have been made with other kind of programmes but EU based ones, hence, he missed a distinct and unambiguous link to EU R&D. Furthermore he said that measuring the impact as presented by Reale can be very complex, as it is always very closely linked to interpretations which can differ a lot from country to country. A third issue Edler addressed was the methodological approach presented in the second speech by Gajdusek; in particular, he found the survey approach "a bit of a danger". Edler suggested that it might be interesting to differentiate between different funders from different countries in order to see national differences in evaluation. The tool presented by Prem very much appealed to Edler; however, what he (and other people in the audience) found a bit dangerous was that if wrong data is used (which may happen due to the rather open access to the tool), all kinds of visualisations and interpretations may come out. In other words, Edler thought that "it is a powerful tool, but it can be misused easily". Prem replied to that concern that he and his colleagues are aware of these dangers. However, he finds that the use and misuse of such a tool is in the end a philosophical question, and the reason why he finds such tools of importance is that there is a lot of interest not only by the European Commission, but also by national policy makers to make use of them. Furthermore he stressed the necessity to critically address issues of misuse, as done by Jakob Edler and people from the audience.

Concerning the fourth speech, Edler liked that it gave a good overview, but he criticised the underlying assumption that through industry involvement everything may become better as too strong in the presentation and quite normative.

SESSION 5: EVALUATING FOR SELECTION – CHALLENGES AND OPPORTUNITIES

NOA BEN-GUR

PRESENTATIONS

Irene DREJER and Poul-H. ANDERSEN, Aalborg University: *Is the tail wagging the dog? An analysis of possible isomorphism effects in innovation project applications*

Peter BIEGELBAUER and Thomas PALFINGER, AIT – Austrian Institute of Technology: Selecting Innovation: *Project Selection Procedures in Research Funding Agencies*

Susanne BÜHRER, Fraunhofer Institute for Systems and Innovation Research: *New modes of stakeholder involvement in ex ante impact assessments*

Kathy WHITELEGG, AIT – Austrian Institute of Technology and Boris KRAGELJ, European Research Executive Council: *Can bibliometric indicators be used to support the European Research Council identify frontier research – and if so how?*

INA DREJER AND POUL-H. ANDERSEN, AALBORG UNIVERSITY:

IS THE TAIL WAGGING THE DOG? AN ANALYSIS OF POSSIBLE ISOMORPHISM EFFECTS IN INNOVATION PROJECT APPLICATIONS

CHALLENGES ADDRESSED

There has been an increasing focus in STI policy regarding goals and effects of research projects. Subsequent to the economic crises of the 00´s is a growing tension in that matter. As risk seems to rise, so does the prerequisite of institution´s demand for getting 'Money´s Worth`.

Nevertheless, there is a paradox which arises: If one can predict the outcome of research, than it is no longer regarded as "innovative". Policy specialists raise the question as to whether the quest for the minimization of risk is promoting safe projects rather than those belonging in the scientific front.

Drejer & Andersen explore the idea of "*mimetic isomorphic effect*" (W.R. Scott, 1995) on project funding application. The research is based on a Danish case study and wishes to address the question to whether the change in guidelines (which shift towards a focus on the ability of applicants to provide explicit documentation and to quantify the foreseen impacts of their research) for funding is effecting the way projects look, in terms of design aim and content.

HYPOTHESIS:

1. The increasing demand for documentation is reflected in mimetic isomorphism. Gaining legitimacy affects the rhetoric and the content of the application.
2. This will lead to more compliance in the expected outcome as the more explicit you are the more vulnerable you are under the evaluation.

Methods and Findings: A buzzword analysis for the following titles: "*Knowledge", "Innovation", "Growth", "Employment" and "Competitiveness"* in 2007 and 2010, following the change in application guidelines.

Findings show a growth in the use of the new rhetoric.

RECOMMENDATIONS/CONCLUSIONS

The study asks further investigation to whether the subjects are also 'walking the walk' and not only changing their rhetoric, as there is no clear sign for value for money. And to see what happens after the funding has been allocated and the work on the project begins.

PETER BIEGELBAUER AND THOMAS PALFINGER, AIT – AUSTRIAN INSTITUTE OF TECHNOLOGY:

SELECTING INNOVATION: PROJECT SELECTION PROCEDURES IN RESEARCH FUNDING AGENCIES

CHALLENGES ADDRESSED

In the last years there has been a growing interest in reviewing elements of funding as part of the social institution of professional scientific work. This work addresses a gap in the literature regarding the procedures of selection themselves.

Biegelbauer and Palfinger investigated nine different European funding agencies in Germany, Austria, Denmark, Sweden Finland and the UK.

The research was based upon conversations and semi-structured interviews with different actors in these agencies. Biegelbauer and Palfinger´s findings show a range of process selection procedures from within the different agencies, let alone amongst themselves.

They also point out to an aptness towards establishing more stable processes of selection in order to reduce operational costs and increasing transparency for both applicants and other actors.

They located the core issue which most agencies spend most of their resources on, and that is the sifting between the "mediocre" proposals.

RECOMMENDATIONS/CONCLUSIONS

Issues found in current procedures deal with a need for a system that reliefs the influence of external factors (such as "the human factor"- personal feelings and former experience with an applicant or an applicant's work) and to supply clear guidelines through which evaluators are able to follow and review the material, as well as to allocate the responsibili-

ty of the final decision making to an overseeing body.

Biegelbauer and Palfinger also recommend to integrate as many external peer reviewers as possible and to set English as a standardized language to assure a sufficient bank of possible personnel. They also suggest creating some kind of peer-network from which experience can be shared.

Finally, Biegelbauer and Palfinger warn from mixing different disciplines in the purpose of evaluation, as parameters and interests may conflict or hinder the decision making.

SUSANNE BÜHRER, FRAUNHOFER INSTITUTE FOR SYSTEMS AND INNOVATION RESEARCH:

NEW MODES OF STAKEHOLDER INVOLVEMENT IN EX-ANTE IMPACT ASSESSMENTS.

CHALLENGES ADDRESSED

STI policy making has shown a change in the last two decades as it wishes to understand different stakeholders´ agendas and involve them in the process of decision making.

In the last years, STI analysts have tried to establish new methods to engage and understand public understanding of STI in order to create an ex-ante assessment for future policy decisions.

Bührer presented the snow-ball online sampling method as a possible tool to reach-out to different groups of stakeholders. She presented the survey conducted for DG RTDI (Directorate-General for Research & Innovation); a semi-delphi structured online survey requesting participants to rate their opinions to statements under 7 different parameters. Although rejected for use, the survey reached a 38% response rate from 300 full unique samples and exceeded the researchers' conservative estimate.

SESSION 6: EVALUATION PRACTICES SCRUTINIZED

FLORIAN BAYER

PRESENTATIONS

Erich PREM, eutema Technology Management GmbH & Co KG: *Evaluation as the construction of policy narratives*

Jürgen STREICHER, Vienna University of Economics and Business: *The Influence of Evaluations on STI Policy Making*

Franz BARJAK, University of Applied Sciences and Arts Northwestern Switzerland FHNW: *Supporting policy learning by means of an evaluation synthesis: findings from a study on Swiss innovation policies*

Wolfgang POLT, Joanneum Research, Kaisa LÄJTEEMÄKI-SMITH and Kimmo HALME, Ramboll Management Consulting: *How to evaluate large-scale 'transformative' STI funding programmes*

In its overall aim Session 6 discussed the usefulness of evaluation practices for policy decisions. The relationships between evaluation processes and policy making were investigated on various levels to contribute to broader questions such as the role and self-understanding of evaluators themselves, the influence of evaluations on concrete decisions, as well as policy narratives, discourses and "policy learning".

KEY CHALLENGES

Erich Prem started off by questioning the self-understanding of evaluators and their profession: by portraying evaluation practices as systematic inquiry to create objective assessments and relying on the notion of detached objectivity produced through instruments, evaluators hide the overall entanglement of evaluation practices with policy narratives. Jürgen Streicher on the other hand directed the attentions towards effects of evaluation exercises and highlighted the tendency towards minor than radical changes and related this to a strong focus on "instruments" in current research on evaluation practices. With a broader approach Franz Barjak discussed results from a synthesis on Swiss evaluation measures over the last fifteen years and scrutinized implementation practices with regard to their policy impact. Finally Wolfgang Polt delivered insights and reflection on the evaluation process within the Finnish Strategic Centres of Science, Technology and Innovation (SHOK) initiative, directing the attention towards the problems when evaluating emerging instruments of a new type with traditional evaluation measures.

KEY DATA

The research presented in the session deals with a vast variety of data: while Prem raised philosophical issues on an epistemic and ontological level, Streicher and Barjak were heavily engaged with data from evaluation processes. The former – following an actor-centred institutional approach – so far engaged in more than 25 in-depth interviews along with document analysis of all kinds relevant to the institutions in question. The latter had to focus on the analysis of evaluation documents "only", mostly due to time restrictions of the project itself. The analysis of Polt et. al. on the other hand is based on first hand insights from three evaluators, discussing an evaluation process they have been engaged with first-hand.

KEY ARGUMENTS

- **Prem**: Evaluation is claiming objectivity, but evaluators always also act as storytellers: they are including values of the overall policy narrative into the evaluation process, while they are creating and shaping the narratives themselves. Therefore evaluators often also act as legitimizers.

- **Streicher**: the presented project sheds new light on evaluation practices, especially the setting up, the process itself and its output and follow-up processes: Since actors in their institutional context have very different agendas when engaging with evaluations, the instrument itself is not that central as usually assumed. While setup and process are usually highly standardized in evaluation procedures, output and follow-ups are very informal and flexible, leading to problems regarding impact and consequences of evaluation practices.

- **Barjak**: Throughout the investigated evaluation practices there appears to be little effort to measure actual goal attainment. There is mostly no reflection of the initial goals and whether they were reached; actors rather quickly go over to looking at the effects of evaluation measures. In other words they try to avoid the focus on failure, in favour of success stories.

- **Polt**: In discussing the SHOK initiative Polt revealed that companies did not buy into the newly developed model, because of too much complexity. Furthermore there was not only reluctance to participate by industry but also by academia. With regard to the expected structural changes the role of incumbents turned out to be too strong and hindered radical shifts. Apart from a few positive results the initiative lead to too little excellence (meaning it did not trigger major breakthroughs in STI) and too little relevance.

KEY RECOMMENDATIONS

Regarding recommendations the central insight of Prem's inquiry is demanding evaluators to make their role as creators of policy narratives explicit. Only by following this rule, ongoing co-productions of policy narratives and evaluation outputs can be reflected. Streicher sug-

gests that more routines and standards could be developed and employed upon output and follow-up processes to better link up evaluation practices and policy. With regard to the big picture – meaning the evaluation of impacts of policy initiatives on society – Polt sees a desperate need for new instruments and approaches. The SHOK initiative furthermore highlights, that policy decisions should stay away from overambitious targets, ill-defined program concepts that are not able to bridge industry and academia as well as programmes that are industry-led only. Furthermore policy making should be cautious against great expectations from large-scale collaborative schemes, clear intervention logic and evaluations based on too little data/information.

OPEN QUESTIONS AND DISCUSSION

The discussion evolved around national differences in how evaluations are integrated into policy: while Austria has a strong focus on evaluation routines, they appear to be less regular and mandatory exercises in Germany. Evaluation practices furthermore need to reflect on different cultures of expressing critique and discussing failures, which also includes very differentiated dealing with the acceptance/role/positioning towards the state as an actor.

The questioning of evaluators as mere legitimizers leaves the controversial question, whether the integration of evaluation outputs and policy narratives is to be considered a problem in the sense of a mere reinvention of the original policy narrative or a process of policy learning.

Last but not least, evaluations have to answer the questions/goals asked at the beginning by the client. Without questions evaluators frame the task themselves, which is contradictory to the dominant self-understanding of the profession: According to this self-understanding questions should be at least negotiated between clients and evaluators. With regard to the first presentation one has to further admit that the description of aims and goals at the very beginning of evaluations is very often based on pre-analysis (e.g. interviews with stakeholders). Therefore evaluators already set up and include part of the overall policy narrative: Is there a lot more interpretation going on than we usually believe?

SESSION 7: EVALUATION OF STI PORTFOLIOS AND POLICY MIXES

SASKIA HABER

PRESENTATIONS

Peter Van den BESSELAAR, VU University Amsterdam Network Institute & Department of Organization Studies. Ulf SANDSTRÖM, Royal Institute of Technology - KTH: *Evaluation at the research systems level: Funding ecologies as policy portfolio*

Christiane KERLEN, Dr Kerlen Evaluation, Christian Von DRACHENFELS, Leo WANGLER and Jan WESSELS, Institut für Innovation und Technik, Volker WIEDMER, Hochschule Magdeburg-Stendal: Portfolio evaluations: *Evaluating policy portfolios and evaluation in a portfolio*

Ly LOOGA, Tallinn University of Technology, Ragnar Nurkse School of Innovation and Governance: *How STI policy instruments affect science and business cooperation in the Estonian ICT sector?*

Edurne MAGRO and James R. WILSOM, Basque Institute of Competitiveness and Deusto Business School, University of Deusto: *Territorial Strategy Evaluation: Beyond Evaluating Policy-Mix*

In this session we were able to listen to four different presentations, each presenting a different project in the field of STI portfolios and the influence of policies.

KEY CHALLENGES AND ARGUMENTS

The first presentation was titled "Evaluating funding modes towards a concept of funding ecologies". Within the presentation the evaluation of funding schemes was a central issue along with questions like the method of comparison, the understanding of dynamics and the question of the optimal system as well as the role of competition. The power point presentation was based on the research results (as far as this was possible since the project is ongoing). A funding system creates a portfolio to everything that's important to the matter or a system. Since the project is ongoing and the data acquired so far are scattered, there is no certainty regarding the outcome, but as far as the evaluation to the present day goes, it seems like the national funding plays a smaller role than the international funding. If funders are small in number, how does this affect the outcome and the size of a possible field? And what about quality?

The second presentation dealt with "Portfolio evaluation: A case study on illustrate evaluation Challenges". It seemed mainly a promotion to raise awareness on the importance of start-ups. For this purpose the data shown gave the impression of the process new companies have to undergo to become successful. The questions that may arise are for example how the contest of choosing could be improved, the role of successful companies and the classification of different roles.

After a short coffee break we heard a presentation on "How STI policy instruments affect science and business cooperation in the Estonian ICT sector"? This was rather interesting since it presented a case study on the national level. One of the challenges this project is facing, next to the EU paradox on the influence on Estonia, is how STI influences political decisions and vice versa. This was pretty interesting since the history is very strongly linked to the development of sciences.

The last presentation was on "Territorial strategy evaluation: Beyond evaluating policy mix". Based on rather recent literature and colourful overwhelming graphics it was mainly stressing the need of strategy evaluation to ensure theoretical policy learning. The field needs to be broadened and policy linked to strategy since policy is just one part of a larger influential group (named government in the presentation).

ADDITIONAL NOTES

After this session I wasn't really sure what I should take with me since I'm not sure if doing evaluations on every single number seems desirable. I understand how this is interesting and important in certain ways but I'm afraid that it will go overboard and we're going to do more evaluations than research.

SESSION 10: EVALUATION OF INTERNATIONAL RTI PROGRAMMES

LEO MATTEO BACHINGER

PRESENTATIONS

Pattharaporn SUNTHARASAJ, National Science and Technology Development Agency of Thailand (NSTDA), Dundar F. KOCAOGLU, Engineering and Technology Management Department, Portland State Unviersity, Oregon: *Evaluating of the International Collaboration in Science and Technology Proposal: How to align the "Curiosity–driven Research" with the "Mission-oriented Goal"*

Isabella E. WAGNER and Stefanie SMOLINER, ZSI – Centre for Social Innovation: *Evaluation of the Austrian bilateral intergovernmental Programme for Science and Technology Cooperation*

Christina SCHUH, Humboldt Foundation: *Expectations on the long-term impact of international research fellowships from a political and an evaluation perspective: challenges and limits to measure side-effects*

International RTI Programmes are diverse and thus difficult/different to evaluate. This session was dedicated to presenting/discussing approaches on how to measure the "success" of RTI-Programmes on transnational scale. Three presenters introduced to the audience approaches/projects dedicated to measure success in terms of input vs. output.

KEY CHALLENGES

One key challenge in the discussions was the "correct" method to measure success. Because of the high differentiation of international RTI-programmes, one of the key challenges is to measure both, intended as well as side effects, of such programmes. Additionally, transparency and steering such programmes was a key difficulty discussed during the session.

KEY DATA

Key data used in the presented projects was on the one hand "hard" output data – publications, contacts, projects, etc. On the other hand, also potential other sources were identified- mainly using qualitative methods to gather data: Experiences, assessments and conclusions by those being funded by the RTI-programmes as well as retracing newly emerged networks as long-term result of international RTI-programmes' input are only some examples.

KEY ARGUMENTS AND RECOMMENDATIONS

To conclude, the presenters agreed that international RTI programmes are difficult to be evaluated. They plead – more or less – for a mixed-methods approach to evaluate such programmes, to consider "hard" as well as "soft", "short-term" and "long-term" output and "intended" as well as "side" effect.

OPEN QUESTIONS AND DISCUSSION

In this session there were only a few questions on general matters at stake rather than follow-up questions for better understanding of the presentations.

CONFERENCE ORGANISATION

The conference "New Horizons / New Challenges: evaluation of STI policies, instruments and organisations" was organised by Austrian Platform Research and Technology Policy Evaluation (fteval), the Manchester Institute of Innovation Research (MIOIR) and L'IFRIS - Institut Francilien Recherche Innovation Societe.

KLAUS SCHUCH

Austrian Platform for Research and Technology Policy Evaluation

JAKOB EDLER

University of Manchester - Manchester Institute of Innovation Research

PHILIPPE LAREDO

Institut Francilien, Recherche, Innovation et Société/ Université Paris-Est

SCIENTIFIC COMMITTEE

Jakob EDLER, Manchester Institute of Innovation Research, University of Manchester

Irwin FELLER, Prof. Em., Economics, Pennsylvania State University

Jordi MOLAS GALLART, Spanish Council for Scientific Research (CSIC)

Dominique GUELLEC, Directorate for Science, Technology and Industry, OECD

Sybille HINZE, ifq-Institute for Research Information and Quality Assurance

Gretchen JORDAN, Center for Innovation/ University of Maryland

Stefan KUHLMANN, Department of Science, Technology, and Policy Studies/ University of Twente

Philippe LAREDO, Institut Francilien, Recherche, Innovation et Société/ Université Paris-Est

Benedetto LEPORI, Centre for Organisational Research/ USI-Università della Svizzera italiana

Maria NEDEVA, Manchester Institute of Innovation Research/ University of Manchester

Wolfgang POLT, Center for Economic and Innovation Research/ Joanneum Research

Klaus SCHUCH, Centre for Social Innovation

Katharina WARTA, Technopolis Austria

ORGANISING COMMITTEE

Jakob EDLER, Manchester Institute of Innovation Research, University of Manchester

Irene DANLER, Austrian Federal Ministry of Science and Research

Abdullah GÖK, Manchester Institute of Innovation Research

Philippe LAREDO, Institut Francilien, Recherche, Innovation et Société/ Université Paris-Est

Sabine MAYER, FFG - The Austrian Research Promotion Agency

Simone MESNER, Austrian Federal Ministry of Science and Research

Rudolf NOVAK, FWF - Austrian Science Fund

Rupert PICHLER, Austrian Federal Ministry for Transport, Innovation and Technology

Sabine POHORYLES-DREXEL, Austrian Federal Ministry of Economy, Family and Youth

Daniela SALHOFER, Platform FTEVAL

Klaus SCHUCH, Centre for Social Innovation

Sonja SHEIKH, Austrian Institute for SME Research

Michael STAMPFER, WWTF - Vienna Science and Technology Fund

Mario STEYER, Austrian Federal Ministry for Transport, Innovation and Technology

Michael STRASSNIG, WWTF - Vienna Science and Technology Fund

Matthias WEBER, Austrian Institute of Technology (AIT) Department Foresight & Policy Development, Austria

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 krho@subnet.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, Technopolis Austria, Vienna Science and Technology Fund (WWTF), The Technology Agency of the City of Vienna (ZIT), 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