44

fteval JOURNAL for Research and Technology Policy Evaluation

ISSUE 44 | SEPTEMBER 2017

JESÚS ALQUÉZAR SABADIE AND CLAIRE KWIATKOWSKI

THE COMMUNITY INNOVATION SURVEY AND THE INNOVATION PERFORMANCE OF ENTERPRISES FUNDED BY EU'S FRAMEWORK PROGRAMMES

YURI SIMACHEV , MIKHAIL KUZYK AND NIKOLAY ZUDIN

ASSESSING THE IMPACT OF PUBLIC FUNDING AND TAX INCENTIVES IN RUSSIA: RECIPIENT ANALYSIS AND ADDITIONALITY EFFECTS EVALUATION

IOANA GALLERON, MICHAEL OCHSNER, JACK SPAAPEN AND GEOFFREY WILLIAMS

VALORIZING SSH RESEARCH: TOWARDS A NEW APPROACH TO EVALUATE SSH RESEARCH' VALUE FOR SOCIETY

FEDERAL MINISTRY FOR SCIENCE, RESEARCH AND ECONOMY

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

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

FEDERAL MINISTRY OF TRANSPORT, INNOVATION AND TECHNOLOGY

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

AQ AUSTRIA

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

AIT - AUSTRIAN INSTITUTE OF TECHNOLOGY

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

AWS - AUSTRIA WIRTSCHAFTSSERVICE GESELLSCHAFT MBH

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

CDG - CHRISTIAN DOPPLER RESEARCH ASSOCIATION

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

CONVELOP COOPERATIVE KNOWLEDGE DESIGN GMBH

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

FFG - AUSTRIAN RESEARCH PROMOTION AGENCY

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

FWF - AUSTRIAN SCIENCE FUND

Sensengasse 1, 1090 Vienna
Dr. Falk Reckling
E: falk.reckling@fwf.ac.at

IHS - INSTITUTE FOR ADVANCED STUDIES

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

JOANNEUM RESEARCH FORSCHUNGSGESELLSCHAFT MBH

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

AUSTRIAN INSTITUTE FOR SME RESEARCH

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

LUDWIG BOLTZMANN GESELLSCHAFT

Nußdorfer Str. 64, 1090 Vienna
Mag. Patrick Lehner
E: patrick.lehner@lbg.ac.at

AUSTRIAN COUNCIL FOR RESEARCH AND TECHNOLOGY DEVELOPMENT

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

ÖAW - AUSTRIAN ACADEMY OF SCIENCE

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

TECHNOLPOLIS AUSTRIA

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

WIFO - AUSTRIAN INSTITUTE OF ECONOMIC RESEARCH

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

WWTF - VIENNA SCIENCE AND TECHNOLOGY FUND

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

WIRTSCHAFTSAGENTUR WIEN. EIN FONDS DER STADT WIEN

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

ZSI - CENTRE FOR SOCIAL INNOVATION

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

INHALT

ISSUE 44 | 2017

EDITORIAL

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

DEAR READERS,

This issue of the fteval Journal for Research and Technology Policy Evaluation contains a number of full papers which were presented at the Open Evaluation conference, organised by the Platform and its partners from the Manchester Institute of Innovation Research (MIoIR) and the Institute for Research and Innovation in Society (IFRIS) in November 2016.

This issue includes several interesting evaluation studies.

Jesús Alquézar Sabadie Claire Kwiatkowski from DG Research and Innovation bring together results of participation of companies in the European Framework Programmes for RTD with results from the Community Innovation Survey (CIS). The analysis of the CIS 2008, 2010 and 2012 demonstrates that innovative enterprises financed by the 7th Framework Programme (FP7) performed significantly better in terms of exploitation of products, services and processes, although there are significant differences by sectors and countries.

The paper from Simachev, Kuzyk and Zudin is politically highly relevant. It aims to assess the additionality effects of direct versus indirect public financial support on companies in Russia. They conclude that tax incentives in Russia almost do not provide significant results in terms of additionality but they also provide arguments for not abandoning tax-based incentives.

Galleron et al. scrutinise the societal value of research in the field of Social Sciences and Humanities. Research evaluation has always been perceived as a difficult area for the SSH for various reasons. Thus, the authors propose an approach looking above all to the combination of performance and valorisation of research in SSH disciplines.

The contribution of Prins and Spaapen is also dealing with research outputs in the field of SSH. They are tracing the impact of some publications from a few institutions in the Netherland operating in the fields of SSH beyond the narrow scientific sphere by applying a Contextual Response Analysis. They found in all cases a variety of stakeholders, also from non-academic fields, interested in the published output.

The paper of Lampert et al. was one of the central conference contributions dealing with openness in evaluation by addressing the quest for suitable indicators to capture and measure open science, a concept which itself is still evolving. They propose indicators for certain dimensions of open science, which, however, are new and not yet gathered/surveyed/evaluated.

We hope that one or the other paper is of interest to you.

Our next conference will take place in Vienna under the auspices of the Austrian EU Council Presidency in early November 2018. The focus of this conference will be on the dominant narrative in research and innovation policy-making in Europe, namely impact of R&D.

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

Klaus Schuch

Stefan Philipp

THE COMMUNITY INNOVATION SURVEY AND THE INNOVATION PERFORMANCE OF ENTERPRISES FUNDED BY EU'S FRAMEWORK PROGRAMMES

JESÚS ALQUÉZAR SABADIE AND CLAIRE KWIATKOWSKI

ABSTRACT

The Horizon 2020 monitoring and evaluation system has been improved in recent years, but there is still a need to further develop the ways to measure innovation outputs, outcomes and impacts. At present, project reporting provides only a few innovation-related indicators. This paper shows that the Eurostat's Community Innovation Survey (CIS) could be a valuable source of information to assess those issues.

The analysis of the CIS 2008, 2010 and 2012 demonstrates that innovative enterprises financed by the 7th Framework Programme (FP7) performed significantly better in terms of exploitation of products, services and processes. The data allow characterising the successful FP7 innovators: large enterprises perform slightly better, and there are significant differences by sector and by country. FP7 funding seems to play a cohesive role amongst countries, as a consequence of cooperative research and innovation activities. Innovative firms supported by FP7 deliver more environmental-friendly innovations and obtain better turnovers from their innovations.

While the CIS could be a useful tool to assess the innovation impacts of the Framework Programmes, there are also some issues to keep in mind. In particular, the design of the questionnaire does not allow for an analysis of a full impact of all FP7 participants: the FP7 had a worldwide participation, while the CIS is limited to the EU respondents. Moreover, confidentiality rules lead to information loses when more than two variables are cross-referenced or when very detailed data (e.g. by NACE beyond one digit) are extracted. Finally, it is important to remember that correlations do not mean causality.

The free and easily accessible CIS data provides a good opportunity to go further in the evaluation of innovation impacts of European framework programmes.

1. BACKGROUND: INNOVATION AND IMPACT IN HORIZON 2020

Innovation is one of the key objectives of the ongoing European Union's *Framework Programme for Research and Innovation*, Horizon 2020 (2014-2020). The EU Regulation 1291/2013 establishes as general objective of Horizon 2020 "(…) to contribute to building a society and an economy based on knowledge and **innovation** across the Union by leveraging additional research, development and **innovation** funding and by contributing to attaining research and development targets (…)" (article 5,§1)[1].

Innovation is not new in the history of the FPs, but it gained a special importance as a response to the global economic crisis that started in 2007[2]. The initial main objective of the *Seventh Framework Programme of the European Community for research, technological development and demonstration activities* (FP7, 2007-2013), the Horizon 2020 predecessor[3], was to implement the European Research Area (ERA). In that sense, FP7 was tasked to: (i) promote transnational cooperation; (ii) promote investigator-driven basic research based on excellence; and (iii) develop the human potential in research and technology, thereby "…encouraging researchers' mobility and career development…"[4]. Concerning innovation, FP7 was initially supposed to complement other EU funding schemes, such as the Competitiveness and Innovation Programme (CIP)[5].

The FP7 orientation changed as a political response to the crisis. In November 2008, the Commission then led by President Durao Barroso launched its Economic Recovery Plan[6]. The document emphasised the need for smart investments, especially on clean technologies, to boost the economy and promote innovation. This narrative was further developed in the Europe 2020 Strategy, which defended a "smart, sustainable

1 Regulation 1291/2013 of 11 December 2013 of the European Parliament and the Council establishing Horizon 2020 - the Framework Programme for Research and Innovation (2014-2020), at: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2013:347:0104:0173:EN:PDF

2 What follows is based on Connolly et al. (2014).

3 Note the different terminology used in the official FP7 and Horizon 2020 names. FP7 was about "research, technological development and demonstration" while Horizon 2020 focuses on "research and innovation". Innovation is for the first time explicitly mentioned in the name of the programme.

4 Decision 1982/2006 of the European Parliament and the Council of 18 December 2006 concerning the Seventh Framework Programme of the European Community for research, technological development and demonstration activities (2007-2013), Preamble, recital (8).

5 Ibid, Preamble, recital (22).

6 Communication from the European Commission (2008) A European Economic Recovery Plan, COM (2008)800 final.

and inclusive growth" concept for Europe. The "smart" component explicitly refers to "knowledge and innovation as drivers of future growth"[7].

Last but not least, the Innovation Union flagship initiative, as part of the Europe 2020 strategy, highlighted the need for action at EU level to develop a strategic approach to research and innovation (R&I). The Commission stated that innovation is "…our best means of successfully tackling major societal challenges, such as climate change, energy and resource scarcity, health and ageing, which are becoming more urgent by the day"[8]. The Commission sometimes presents Horizon 2020 as "the financial instrument implementing the Innovation Union, a Europe 2020 flagship initiative aimed at securing Europe's global competitiveness"[9].

This historical background explains the strong focus of the current Framework Programme, Horizon 2020, on innovation. Innovation is seen as a solution to address the economic crisis, while tackling major societal challenges. This idea is at the core of the rationale of Horizon 2020.

The Commission has the legal obligation to evaluate the results of the Framework Programmes. However, assessing the impact of R&I is more important than ever in Horizon 2020, which regulation contains many references like "achieve maximum impact", "achieving the greatest possible impact" or "maximise impact". This is further developed in the *Council Decision establishing the specific programme implementing Horizon 2020* ("Specific Programme")[10], which for the first time includes an Annex with "performance indicators". Some few relate to innovation:

- Patent applications and patents awarded in Future and Enabling Technologies, in the different enabling and industrial technologies and in the various societal challenges (i.e. under the three Horizon 2020 main priorities: "Excellent Science", "Industrial Leadership" and "Societal Challenges").
- Share of participating firms introducing innovations new to the company or the market (covering the period of the project plus three years), under the priority "Industrial Leadership".
- Number of prototypes and testing activities, under the priority "Societal Challenges".

For the Horizon 2020 monitoring and evaluation, the compulsory performance indicators represent an improvement compared with previous Framework Programmes. Positive developments have occurred in recent years, for example through the creation of RESPIR, the Commission database on R&I projects' outputs[11]. However, there is a constant need to enhance the evaluation and monitoring system of the Framework Programmes. The High Level Expert Group for the Ex Post Evaluation of FP7 observed and recommended that „(…) evaluation activities have been considered as routine activities in recent years (…). Considering that the Framework Programme have consistently been the third largest budget of the European Union, a strategic and professional monitoring and evaluation system is required that increases transparency and serves as a comprehensive and trusted source of evidence-based decision making" (Martinuzzi et al. 2015, p.9).

This diagnosis is not new for the Commission services dealing with monitoring and evaluation of the Framework Programmes. Other previous evaluation exercises reached similar conclusions. For instance, the *Ex Post Evaluation of FP7- Cooperation Theme: Environment (including Climate Change)* recommended the Commission to enhance its monitoring system, especially in the areas of innovation and policy use of results. The authors of this assessment said that "the Commission [should be able] to identify innovative projects with potential societal impacts, as well as their strengths and weaknesses, to provide further support (if needed) and facilitate networking with complementary projects, and dissemination. For innovation issues, the monitoring system should rely on a set of smart indicators (…) and on insights from Project Officers" (Connolly et al. 2014, p.79).

Nevertheless, the management of Horizon 2020 is confronted to a paradox. Innovation is one of the core issues that the programme should address (it was indeed judged as the key European weakness to push for growth and to address societal challenges in the *Horizon 2020 Ex Ante Impact Assessment*[12]), but it remains weakly monitored. It would be very relevant for the Commission to explore new tools to follow-up systematically and comprehensively the innovation results of projects and their impacts. There is still a need to collect basic information on innovation outputs and outcomes, like Technology Readiness Level (TRL) attained, barriers encountered to commercialise or exploit results, health/energy/resource efficiency/climate impacts of innovations (e.g. reduction of emissions, saving of energy or raw materials), commercialisation data or further investments committed. It is increasingly necessary to measure the creation of economic value and impacts of projects, in order to answer the following questions: What are the economic returns of participating in Horizon 2020? Is participation paying off economically?

How can the Commission assess basic economic (and environmental) impacts of Horizon 2020? This question is now politically critical, because of the strong focus of the current President Juncker Commission on growth, jobs and investment[13]. In this context, EU-funded R&I must demonstrate its impact and contribution to such economic goals, in a time when austerity measures strongly affect research funding in several European countries.

7 Communication from the European Commission (2010) Europe 2020: A strategy for smart, sustainable and inclusive growth, COM (2010)2020 final.
8 Communication from the European Commission (2010) Europe 2020 Flagship Initiative Innovation Union, COM (2010) 546 final.
9 See the Horizon 2020 website, at: https://ec.europa.eu/programmes/horizon2020/en/what-horizon-2020
10 Council Decision of 3 December 2013 establishing the specific programme implementing Horizon 2020 - the Framework Programme for Research and Innovation (2014-2020).
11 RESPIR stands for RESearch Performance and Impact Reporting tool. In includes data on publications, Intellectual Property Rights (IPRs), dissemination activities and workforce statistics, based on projects' reporting. Before the last years of FP7, those data were collected, but never handled and aggregated in a common and harmonise database. Evaluation studies used to launch new surveys to collect information on concrete outputs from projects.
12 Commission Staff Working Document accompanying the Communication from the Commission ‚Horizon 2020 - The Framework Programme for Research and Innovation', SEC(2011)1487 final. At: http://ec.europa.eu/research/horizon2020/pdf/proposals/horizon_2020_impact_assessment_report.pdf
13 See: https://ec.europa.eu/priorities/index_en

2. THE COMMUNITY INNOVATION SURVEY AND THE FRAMEWORK PROGRAMMES

The main source of information on R&I outputs and outcomes is projects' reporting. The simplification principle guides the whole Horizon 2020 implementation. It implies that the extent and content of reporting is rather limited under this Framework Programme. In its response to the recommendations of the High Level Expert Group for the Ex Post Evaluation of FP7, the Commission commits to "establish data links with external databases to complete and improve the quality of data sets"[14]. Solutions and alternative data sources must be therefore explored and used to assess the innovation results and impacts of Horizon 2020 projects.

Eurostat launched its first Community Innovation Survey (CIS) in 1992. This bi-annual large scale survey provides harmonised data on enterprises' innovation activities and results by sector, size of company, type of innovation and the various stages of the innovation process: objectives, sources of information, investments, public funding, etc. The CIS is carried out in all Member States and other associated countries, but not all of them allow accessing to the raw data through Eurostat. For example, the CIS 2012 made data accessible for 13 EU Member States plus Norway[5], with 143,669 enterprises covered.

The CIS questionnaire includes an item that refers to funding from the Framework Programme[16]:

During the three years 2010 to 2012, did your enterprise receive any public financial support for innovation activities from the following levels of government? (Include financial support via tax credits or deductions, grants, subsidised loans, and loan guarantees. Exclude research and other innovation activities conducted entirely for the public sector under contract).

- *Local or regional authorities – Yes/No*
- *Central government (including central government agencies or ministries) – Yes/No*
- *The European Union (EU) – Yes/No*
- *If yes, did your enterprise participate in the EU 7th Framework Programme for Research and Technical Development? – Yes/No*

The last question (indicator FUNRTD in Eurostat's nomenclature) is very relevant for evaluating FP7 outcomes and impacts. It allows identifying in an aggregated way, within the CIS respondents, enterprises that received FP7support.This makes it possible to perform a counter-factual analysis, comparing results of companies that received FP7 funding with those that did not (but could benefit from other financial support, local, regional or European). The CIS data also permit to understand which factors and barriers influence innovation outcomes, both for FP7-funded enterprises and overall.

Despite a large scientific literature that exploits the CIS, especially at national level, the Commission services have rarely used it to evaluate the EU's Framework Programmes. The European R&I policy community largely relies on traditional indicators, like publications and patents, insufficient to capture innovation. Indeed, academic analysis of the CIS has had little impact on the European innovation policy (Arundel 2007). There are however some exceptions. In 2009, a PRO-INNO report combined the analysis of CIS with ad-hoc surveys and case studies to conclude that "the Framework Programme attracts the highly innovative companies and research institutions in Europe". The participants were more R&D intensive, more networked and more internationalised than the average. They obtained higher returns on innovative sales (Fisher, Polt & Vonortas 2009, p.7-8). The study referred to FP4 (1994-1998) and FP5 (1998-2002). Muldur et al. (2006) reached similar conclusions in 2006, using the CIS 3 (1998-2000).In 2013, a Science-Metrix study on Small and Medium Enterprises (SMEs) innovation performance used the CIS 2010 to design and test an ad-hoc survey questionnaire (Hassan et al. 2013).

Using the CIS for evaluation purposes presents nevertheless some caveats:

- **Geographical coverage:** Not all EU Member States make their CIS raw data available, while the FPs are open to the world. All countries can participate in Horizon 2020, with different status and under different conditions: Member States, candidate countries, associated countries, developed or emerging third countries, developing third countries. Therefore the available CIS raw data give an incomplete picture of the innovative firms involved in the Framework Programmes.
- **Timeframe:** The CIS is a bi-annual survey[17], which dates do not necessarily coincide with the starting and ending years of the FPs. For example, FP7 ran from 2007 to 2013. The CIS 2008 covered years 2006 to 2008; therefore its FUNRTD variable referred to both FP6 and FP7. Similarly, the CIS 2012 went from 2010 to 2012 and thus lets out the last year of FP7, 2013. It is important to point out, indeed, that almost half of the FP7 projects were still running when the programme was replaced by Horizon 2020 in 2014. These projects will provide their main outcomes and impacts long time after 2013.
- **Issues related to the questionnaire design**: There are typically five types of innovation: on product, service, process, organisational and marketing. Due to its position in the questionnaire, the item on FP7 support just covers the three first categories. In addition, the FP7-related question is filtered. Only enterprises that declared having introduced during the last three years any product, service or process innovation or, at least, having been involved on any innovation activity abandoned or still ongoing, were allowed to answer the question on FP7 funding. This means that only innovative companies supported by FP7 can be analysed; FP7-funded enterprises not involved in innovations cannot be identified. Even if we could expect that private for profit organisations' main motivation to participate in FPs may

14 Communication from the European Commission(2016) Communication on the Response to the Report of the High Level Expert Group on the Ex Post Evaluation of the Seventh Framework Programme, COM (2016)5 final, p.9.

15 See: http://ec.europa.eu/eurostat/web/microdata/community-innovation-survey . The CIS 2012 methodology is explained in detail at: http://ec.europa.eu/eurostat/cache/metadata/en/inn_cis8_esms.htm

16 Question 5.3 in CIS 2012.

17 The CIS is bi-annual since 2007. The first four editions covered three-years-time intervals. The story of the CIS can be read in Hassan, E. et al. (2013), from p. 12.

be precisely innovation, we cannot assume a priori that the CIS sub-sample of FP7-financed enterprises is a representative sub-sample of all companies that benefit from FP7. The CIS allows to analyse innovative enterprises supported by FP7, but not necessarily all enterprises supported by FP7.

- **Anonymisation of respondents**: The CIS must comply with strict confidentiality rules. This is an issue for researchers and analysts, who cannot merge the data with other sources (e.g. internal database of the European Commission on FP projects, CORDA) and cannot track firms over time. Thus it is difficult to analyse the circular link between policy, R&D, innovation and performance (Mazzanti et al. 2016).
- **Problems of the eco-innovation module**: The CIS 2008 included a voluntary eco-innovation module, whose next edition is foreseen in CIS 2014. It provides useful information about environment-friendly innovations introduced by enterprises and why they were implemented. The purposes of the eco-innovation are very focused on policy aspects: regulations, taxes, public support or voluntary codes. There is just one item on economic aspects ("current or expected market demand from your customers for environmental innovation"). Although this information is precious, it hinders some relevant aspects like the economic mitivations (cost reduction, productivity growth, competitive advantage, etc.). Ethical considerations, like companies that develop environment-friendly innovation for ethical reasons, are neither taken into account. Last but not least, the filters of the questionnaire allow analysing a sub-sample of eco-innovative enterprises, but nothing is known about how regulation influences the behaviours of non-eco-innovative firms (Mazzanti et al. 2016).

Despite these caveats, the CIS is a very relevant source of information to assess and analyse the innovation results of the Framework Programmes. This is the purpose of this paper. The authors use data from the CIS 2008, 2010 and 2012[18] and try to answer to the following questions: Do enterprises supported by FP7 perform better than the average? Are there significant differences by country, sector, size of enterprise, source of financing, etc.? What are the economic returns of exploited innovations? The analysis focuses on FP7 funding but aims at demonstrating the opportunity for the Commission services of exploiting the CIS data systematically to assess innovation results and impacts of the Framework Programmes. From the European Commission perspective, the final goal of the analysis is to extract concrete and operational lessons from FP7 which can be used for the Interim Evaluation of Horizon 2020.

The paper looks also at the results of the Eco-Innovation module proposed at the CIS 2008 and links its results with the general innovation trends. This is particularly relevant in the context of the Circular Economy strategy of the EU, which must be monitored too[19].

3. DO INNOVATIVE ENTERPRISES SUPPORTED BY FP7 PERFORM BETTER IN TERMS OF ECONOMIC OUTPUTS?

We consider that enterprises perform better in terms of innovation when they:

- Introduce a new or significantly improved product to the market before their competitors do (variable NEWMKT).
- Introduce a new or significantly improved product to the firm, which was already available from competitors in the reference market (variable NEWFRM).
- Introduce new or significantly improved process innovations (i.e. methods of manufacturing or producing goods and services; logistics, delivery or distribution methods for inputs, goods or services; or supporting activities for processes, such as maintenance systems or operations for purchasing, accounting or computing), new to the market (variable INPSNM).

Therefore this analysis uses the exploitation of innovations as a performance indicator. It does not look at the advancement in the innovation process, i.e. from a Technology Readiness Level (TRL) to another.

Table 1 compares the innovation performance of innovative enterprises that benefitted from FP7 funding with those that did not.

Table 1: Innovation performance: Firms supported by FP7 vs. not supported

CIS 2008 (2006-2008)				
	Supported by FP7	**Non-supported by FP7**	**Significance Chi-square**	**Phi coefficient**
New to the market product or service innovations (NEWMKT)	1,132 73.36%	13,376 42,67%	<0.0001	0.13
New to the firm product or service innovations (NEWFRM)	1,082 71.14%	17,554 56.02%	<0.0001	0.064
New to the market process innovations (INPSNM)	357 39.23%	3,471 19.8%	<0.0001	0.106
CIS 2010 (2008-2010)				
New to the market product or service innovations (NEWMKT)	1.076 79.79%	11,575 31.59%	<0.0001	0.186
New to the firm product or service innovations (NEWFRM)	917 70.38%	15,299 41.72%	<0.0001	0.106
New to the market process innovations (INPSNM)	362 49.05%	3,048 12.83%	<0.0001	0.203

18 Their sample size and geographical coverage is presented in Annex 1.

19 Communication from the European Commission (2015) Closing the loop – An European action plan for the Circular Economy, COM(2015) 614 final

CIS 2012 (2010-2012)				
New to the market product or service innovations (NEWMKT)	1.191 78.51%	10,144 43.38%	<0.0001	0.169
New to the firm product or service innovations (NEWFRM)	943 66.74%	13,821 59.38%	<0.0001	0.035
New to the market process innovations (INPSNM)	378 42.81%	2,732 18.83%	<0.0001	0.156

FP7-funded innovative enterprises perform significantly better than those not supported. Between 2006 and 2012, more than 70% of the firms that benefitted from FP7 funding introduced new products to the market, while others remained under 45%. The difference is less pronounced – but still very strong – when referring to products new to the firm only, while new to the market process innovations present lower figures (below 50% in all CIS 2008, 2010 and 2012) that anyway double those of enterprises not supported by FP7.

Of course, the significant correlations between FP7 participation and innovation performance do not necessarily mean causality. It could be assumed that the Framework Programmes attract R&I-intensive organisations, which are expected to be more innovative than the average.

Indeed, FP7 was a R&I programme focused on excellence, with a very low success rate of applications (18.7%)[20]. Beneficiaries needed to have very strong capacities to be selected by independent evaluators. Amongst the main FP7 beneficiaries we find the biggest European R&I organisations, such as the Centre National de la Recherche Scientifique (CNRS, France), Fraunhofer (Germany), the universities of Cambridge and Oxford, the Commissariat à l'Energie Atomique et aux Energies Alternatives (CEA, France) or the German Max Plank Institute. Amongst private for profit organisations, the ranking (by EU contribution received) is led by Geant Limited (UK, on water transport), SAP (Germany, ICT), Thales, Siemens, etc. The top European firms in terms of R&D investment appear also in the list of FP7 participants[21].

Therefore Table 1 does not allow to conclude that firms financed by FP7 obtain better innovation results because of their participation in the programme. However, FP7 was at least likely to attract the most R&I intensive enterprises, which in turn improve their capacities thanks to collaboration in R&I at international level. The logic is likely to be circular, not linear.

The main differences between FP7-funded firms and enterprises not supported by the EU's Framework Programme appear for new to the market product and services innovations, and for new to the market processes. This indicates that FP7 led primarily to the development and implementation of novel products, services and processes and not to replicate or improve those that were already in the market. The opposite could be considered as a failure for a R&I programme.

The results presented in Table 1 hide differences by sector, size of enterprise and country. Not all enterprises that participate in FP7 obtain similar results.

3.1. INNOVATION PERFORMANCE, BY SIZE OF ENTERPRISE

The Horizon 2020 Regulation considers that "SMEs [micro, Small and Medium-sized Enterprises] constitute a significant source of innovation, growth and jobs in Europe"[22] and provides different specific instruments to push for innovation in SMEs. It is therefore very relevant to breakdown the innovation results of European enterprises by size, comparing those funded by FP7 with other firms.

This analysis distinguishes micro, small and medium enterprises based on the number of employees and the turnover criteria: Micro SMEs are those with less than 10 employees and €2,000,000 turnover or less; Small SMEs employ less than 50 people and their turnover is lower or equal to €10,000,000; Medium SMEs have less than 250 employees and no more than €50,000,000 of turnover. Beyond these limits, enterprises are considered "large". These categories follow the European Commission's definition, which however adds a criterion based on the balance sheet[23], not considered here because not covered by the CIS.

Table 2 shows the breakdown of enterprises that received support from FP7 and declared being involved in innovation activities, by size. As a reference, 55% of private-for profit organisations that participated in FP7 were SMEs.

Table 2: FP7-funded enterprises with innovation activities, by size (%)

	Micro	Small	Medium	Large	N
CIS 2008	8.58%	26.47%	30.62%	34.32%	1,783
CIS 2010	7.80%	29.40%	26.15%	36.65%	1,629
CIS 2012	6.47%	28.83%	26.62%	38.07%	1,807

Does the size of enterprises matter in terms of innovation performance (i.e. exploitation of innovations)? – Annex 2 presents the contingency tables that help to answer this question, for CIS 2008, CIS 2010 and CIS 2012. They focus on new to the market and new to the firm products and services (variables NEWMKT and NEWFRM). Tables for the variable INPSNM ("new to the market process innovations") are not used, because of the large number of data missing due to the CIS' confidentiality rules.

The CIS data show that large innovative companies perform better in introducing new to the market products or services (variable NEWMKT). This trend appears in all three CIS editions, for both FP7-funded and not funded enterprises, except in CIS 2010 for FP7-supported enterprises. For instance, in CIS 2008, 79% of large innovative enterprises funded by FP7 introduced a new product or service to the market, while the figures for SMEs were around 70%. In 2012, the differences were much lower (82.4% versus 79.6% for micro SMEs) and even statistically insignificant in 2010.

Amongst FP7-funded enterprises, most new to the market innovations come from large companies, while small SMEs present the biggest figure for non-FP7-financed firms. This is not surprising. Small SMEs constitute the largest share of non-FP7-financed innovative enterprises,

20 Source: CORDA database. The four main FP7 specific programmes ("Cooperation", "Ideas", "Capacities" and "People") plus Euratom received 135,799 proposals and only 25,363 were selected for funding.

21 See the main R&D investors in Europe at the annual EU Industrial Scoreboard reports prepared by the European Commission, Joint Research Centre, at: http://iri.jrc.ec.europa.eu/scoreboard.html

22 Regulation 1291/2013, recital (34).

23 See: http://ec.europa.eu/growth/smes/business-friendly-environment/sme-definition/index_en.htm

while large companies are the most numerous amongst firms participating in the programme.

Concerning new to the firm innovations, the best performing enterprises supported by FP7 are large and micro ones. Differences by size are nevertheless minor.

The size of the enterprises influences different levels of innovation outputs, but it is not a critical factor: Phi and contingency coefficients are low, between 4% and 16%. Differences are therefore small and sometimes even statistically insignificant. In any case, large firms tend to perform slightly better in terms of introducing products and services new to the market. Companies supported by FP7 obtain always better innovation outcomes, irrespective of the size of the enterprise.

3.2. INNOVATION PERFORMANCE, BY COUNTRY

Some countries perform significantly better than others in terms of exploitation of innovations. Indeed, the variable "country" influences more the introduction of new products, services or processes to the market or the firm than the size of enterprises. Table 3 presents the contingency and Phi coefficients of cross-tables between countries and NEWMRKT, NEWFRM and INPSNM variables. In all cases, the correlation is statistically significant and its degree of association much higher than those obtained for size of enterprise.

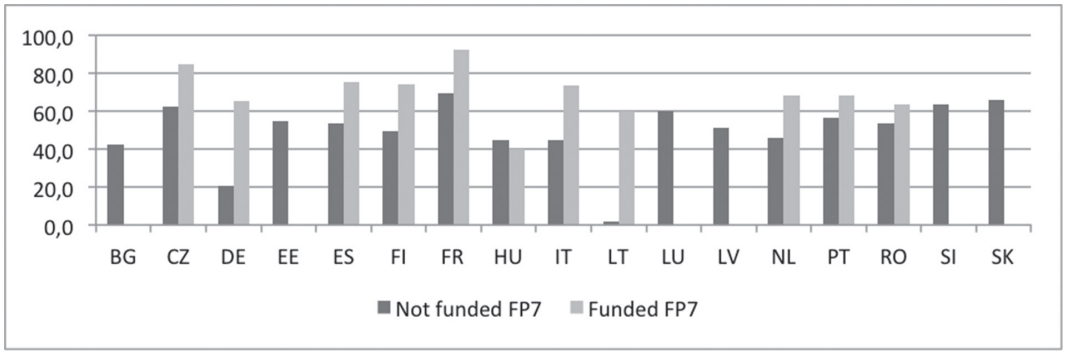
Table 3: Degree of association country-exploitation of innovation variables

	New to the market product or service innovations (NEWMKT)	New to the firm product or service innovations (NEWFRM)	New to the market process innovations (INPSNM)
CIS 2008	Sign. Chi-square <0.0001 Phi coeff.: 40.37% Contingency coeff.: 37.43%	Sign. Chi-square<0.0001 Phi coeff.: 53.97% Contingency coeff.: 47.49%	Sign. Chi-square<0.0001 Phi coeff.: 69.79% Contingency coeff.: 57.23%
CIS 2010	Sign. Chi-square <0.0001 Phi coeff.: 43.95% Contingency coeff.: 40.23%	Sign. Chi-square<0.0001 Phi coeff.: 56.37% Contingency coeff.: 49.1%	Sign. Chi-square<0.0001 Phi coeff.: 48.57% Contingency coeff.: 43.69%
CIS 2012	Sign. Chi-square <0.0001 Phi coeff.: 46.45% Contingency coeff.: 42.27%	Sign. Chi-square<0.0001 Phi coeff.: 52.51% Contingency coeff.: 46.49%	Sign. Chi-square<0.0001 Phi coeff.: 48.78% Contingency coeff.: 43.84%

Figure 1 illustrates the difference performance of countries, either when "their" enterprises receive FP7 funding or not. Unfortunately, due to confidentiality rules, the contingency tables between implemented innovations and countries controlled by FP7 funding cover only few countries: 11 in the CIS 2008, 7 in CIS 2010 and 9 in CIS 2012, when referring to new to the market products and services (NEWMKT). Those countries represent, however, 62% of the FP7 funding in CIS 2008, 52% in CIS 2010 and 58% in CIS 2012.

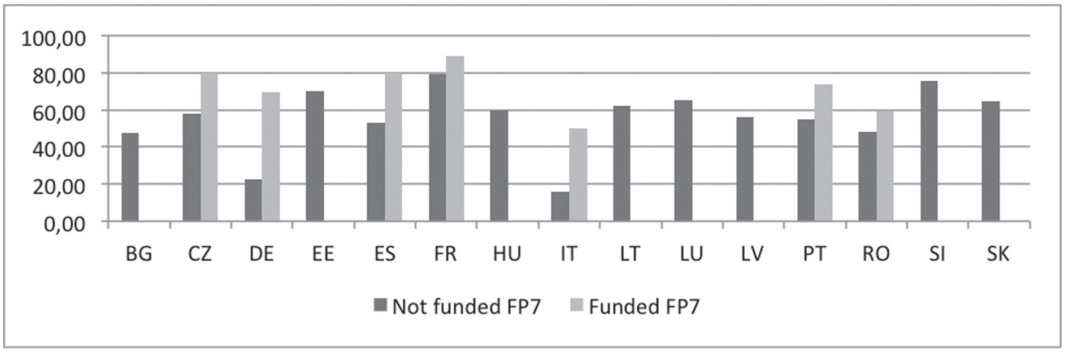
Figure 1: New to the market product innovations, by country

CIS 2008



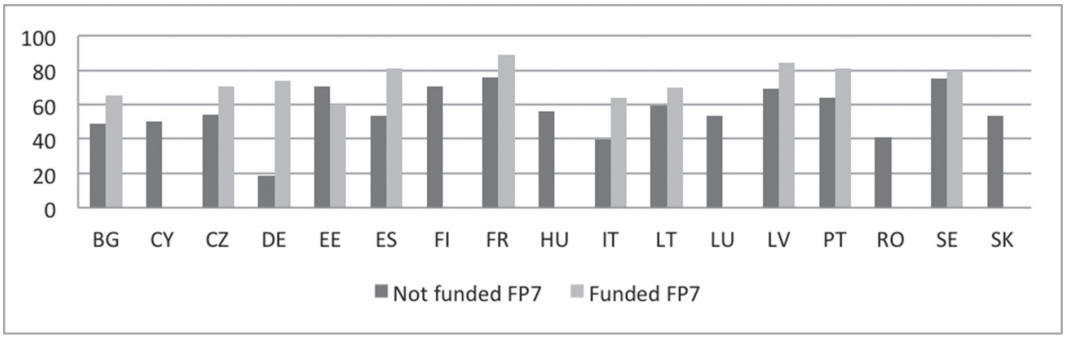
N: 13,277 innovative enterprises not funded by FP7 (42.6% of the total) and 1,047 enterprises funded by FP7 (73.9% of the total)

CIS 2010



N: 10,194 innovative enterprises not funded by FP7 (43.5% of the total) and 833 enterprises funded by FP7 (77.3% of the total)

CIS 2012



N: 11,492 innovative enterprises not funded by FP7 (31.5% of the total) and 996 enterprises funded by FP7 (77.6% of the total)

Companies located in EU Member States like the Czech Republic, Spain, France or Portugal perform best in terms of introduction of new products or services to the market. Participation in FP7 has a statistically significant positive impact in all countries and all CIS editions.

Germany is an interesting case. In general, a low proportion of its firms say that they introduced new products to the market (24.7% in CIS 2008, 26.6% in CIS 2010 and 22.5% in CIS 2012). However, innovative firms from Germany supported by FP7 showed considerably higher figures (65.7%, 69.7% and 73.7% respectively). These figures are two to three times higher than of those companies that did not receive FP7 support. In most countries, FP7 participation increases the performance of companies by 25% to 50%.

It is also interesting to note that differences by countries are lower when referring to enterprises funded by FP7. Phi and contingency coefficients, even if statistically significant, are much lower: 22%-23% in CIS 2008 and CIS 2010, and even slightly under 20% in CIS 2012. This may indicate that, in terms of introduction of new products or services to the market, FP7 also plays a cohesive role for innovative enterprises, as a consequence of collaborative R&I activities.

These trends are not so evident when looking at the impact of FP7 participation on new to the firm innovations by country. FP7 involvement does not always make a difference for this sort of innovations. Again, the data suggest that innovative enterprises involved in FP7 tend to focus on new to the market novelties.

Due to confidentiality rules too many data are lost to analyse the variable "new to the market processes" (INPSNM).

3.3. INNOVATION PERFORMANCE, BY SECTOR (NACE 1 DIGIT)[24]

Table 4 shows that the Manufacturing sector (NACE C) provides the majority of new to the market product innovations of the European economy, followed by Information and Communication (NACE J, around 10-14%) and by "Wholesale, retail and repair of vehicles" and "Professional, scientific, technical activities" (NACE G and M respectively, slightly less than 10% each).The innovation performance of sectors is uneven, with statistically significant differences between them (Phi Coefficient and Contingency coefficient between 23% and 32%, depending on the year).

Within innovative enterprises supported by FP7, NACE C, NACE M and NACE J cover alone 90% of the new products to the market. The manufacturing sector represents a share of more than 50%, Scientific and technical services around 25% and ICT 12-14%.

Even more interesting is to observe the gap in terms of innovation performance between companies not funded by FP7 and those that were. In the CIS 2008, 73.2% of the innovative manufacturing enterprises supported by FP7 introduced a product new to the market. This percentage grew to 79.4% in the CIS 2012. The figures are between 57% and 84% higher than amongst enterprises that did not benefit from FP7 support.

In other sectors, the positive differences are even higher: +78% to +113% in the Scientific and Technical services sector, or even more than +200% in NACE D, "Electricity, gas, air conditioned supply".

ICT enterprises were the most successful in terms of introduction of new products to the market. In CIS 2012, for instance, 56.3% of the ICT enterprises not supported by FP7 introduced a new product to the market, and 76.5% of those supported. The latter figure was even higher than 80% in previous CIS editions.

Similar trends appear for new to the firm innovations. Enterprises funded by FP7 perform much better than those not funded, in all sectors for which data are available.

24 See Eurostat (2008) p. 57.

Table 4: New to the market product innovation, by sector

	NACE A	NACE B	NACE C	NACE D	NACE E	NACE F	NACE G	NACE H	NACE I	NACE J	NACE K	NACE L	NACE M	NACE N	NACE P	NACE Q	NACE R	NACE S
CIS 2008																		
Whole economy (% by row)	0.45	0.41	56.21	0.46	1.31	3.28	9.68	2.77	0.92	10.88	4.16	0.26	7.42	1.1	0.1	0.34		0.19
Funded FP7 (% by row)			50.97	1.5		1.33	2.92	1.33		13.07		0	26.06					
Not funded by FP7	49.75	17.6	46.76	16.31	21.7	31.57	41.87	24.53	30.89	54.51	44.08	28.57	41.82	26.05	40.68	36.41	45.45	51.19
Funded by FP7			73.22	58.62		62.5	63.46	44.12		80.43		0	78.67					
CIS 2010																		
Whole economy (% by row)	0.46	0.42	57.29	0.4	1.4	2.94	8.67	2.36	0.16	12.11	3.75	0.1	7.94	0.71	0.22	0.68	0.14	0.26
Funded FP7 (% by row)			53.85	1.78				2.44	0	13.95			22.51					
Not funded by FP7	48.62	15.03	43.03	11.19	15.64	8.48	16.64	14.39	27.14	51.28	29.23	52.38	39.09	20.5	55.32	37.09	41.03	53.57
Funded by FP7			78.95	63.33				65		81.42			83.33					
CIS 2012																		
Whole economy (% by row)	0.2	0.48	54.22	0.63	1.2	2.01	9.78	2.4	0.17	14.2	3.72		9.73	0.41	0.11	0.5		0.13
Funded by FP7 (% by row)			53.57	1.85	0.92		2.43		0	11.75		0	25.44					
Not funded byFP7	48.61	28.1	47.1	18.65	22.45	31.56	41.94	22.23	42.86	56.28	37.33		46.54	15.58	44.44	40	40	
Funded by FP7			79.35	61.11	50		72.5			76.5			83.01					

3.4. ECO-INNOVATION

CIS 2008 included a module on "innovation with environmental benefits". According to the survey, an environmental innovation is a new or significantly improved product (good or service), process, organisational method or marketing method that creates environmental benefits compared to alternatives. The definition includes also the following precisions:

- The environmental benefits can be the primary objective of the innovation or the result of other innovation objectives.
- The environmental benefits of an innovation can occur during the production of a good or service, or during the after sales use of a good or service by the end user.

This definition is voluntarily broad and highlights the multidimensional and systemic aspects of eco-innovation which can be seen as "a series of connected changes improving or creating novel functional systems that reduce use of natural resources and decreases the release of harmful substances across the whole life cycle" (European Commission 2015, p.11, based on EIO 2013). These specific characters of eco-innovation make it usually difficult to measure it notably through macro-indicators which tend to be mainly sectoral. Micro-data such as those made available through the CIS 2008 are therefore essential to facilitate the assessment of the private R&I environmental performance in Europe. They offer tailored information which could be used notably in the context of Horizon 2020 and the European Commission's Circular Economy package which was published in 2015.

Horizon 2020 is organised around "three mutually reinforcing priorities: (a) Excellent Science; (b) Industrial leadership; (c) Societal Challenges"[25]. Societal Challenges include environmental and climate-related actions, in line with the Europe 2020 strategy for a "smart, sustainable and inclusive growth", including sustainable agriculture, clean and efficient energy, green transport, climate action and resource efficiency, etc. Climate action and sustainability are overarching principles of Horizon

25 Horizon 2020 Regulation, op.cit, art. 5, §2.

2020. It is indeed expected "(…) that at least 60 % of the overall Horizon 2020 budget should be related to sustainable development. It is also expected that climate-related expenditure should exceed 35 % of the overall Horizon 2020 budget, including mutually compatible measures improving resource efficiency"[26].

The Commission services monitor the funding of sustainable development and climate change, but what about the results of these investments? The CIS allows quantifying innovations with environmental benefits, such as materials or energy savings, reduced CO2 footprints, waste or water efficiency, etc. Of course, CIS 2008 refers to FP6/FP7 instead of Horizon 2020, but the data can at least give a flavour of the environmental impact of innovations supported by the FPs. It allows answering the following question: To what extent are innovations funded by FPs environmental-friendly?

To facilitate the analysis, the authors have created a new variable ECOTOT combining the nine questions referring to environmental benefits of innovations available in the CIS 2008 questionnaire (Q10.1). If respondents indicated at least one of these benefits proposed, ECOTOT would be positive. In other words, ECOTOT means "at least one environmental benefit of the innovation, either from the production of goods or services, or from the after sales use by the end user". This variable is then crossed by FUNRTD, as presented in Table 5.

Table 5: Environmental benefits of innovation (ECOTOT), by FP6/FP7 funding

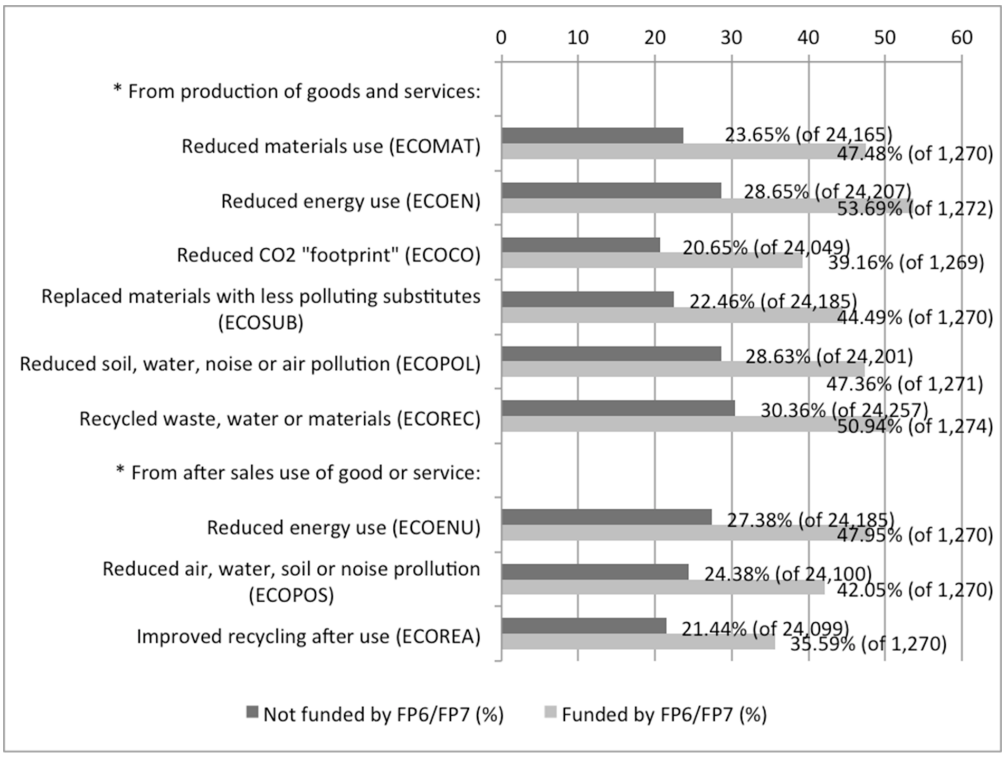
	No environmental benefit	Environmental benefit	N
Not funded by FP6/FP7 (% by row)	80.12%	19.88%	65,180
Funded by FP6/FP7 (% by row)	43.63%	56.37%	1,783
Total (% by row)	79.15%	20.85%	66,963

Source: CIS 2008

The relationship is statistically significant (Phi coefficient = 14.5%, Contingency coefficient = 14.3%). Innovative companies supported by FP use to deliver more environmental-friendly products and services. They introduced almost three times more eco-innovations than enterprises that did not benefit from FP's support.

Figure 2 shows which sorts of eco-innovations are the most frequent ones. It is worth noting that multiple answers were possible: an eco-innovation can have multiple kinds of benefits on environment.

Figure 2: Eco-innovations by type and by FP6/FP7 funding



26 Ibid., recital (10).

As regards the type of eco-innovation, the ranking is different amongst enterprises not funded by FP6/FP7 and amongst those that received the EU's R&D support. The former tended to introduce innovations aiming at:

- recycling waste, water or materials (30.4%),
- reducing energy and soil, water, noise or air pollution during the production (28.7% - 28.6%),
- saving energy during the use of the good or service (27.4%).

Innovative enterprises funded by FP6/FP7 tended to introduce eco-innovations aiming at:

- saving energy during the production (53.7%),
- recycling of waste, mater or material during the production (50.9%), and
- saving energy during the use (47.95%).

Overall, environmental benefits during the production (i.e. to reduce costs) are dominant over those focused on the use of the final product or service (18.4%/52.9% versus 14%/43.1% respectively).

Drivers of eco-innovation can be assessed trough question Q10.2. The main reason that motivates eco-innovation is the existence of regulations and taxes (24.7%/44.7% for FP7 non-funded and funded firms respectively). Surprisingly, grants, subsidies or other financial incentives are the less often quoted factors. This is also true for FP6/FP7-funded companies (10.4%/22.7%). Voluntary codes or agreements, future regulations or market demand are mentioned by similar percentages of respondents (17-19%/38-39%).These observations confirm the results of the economic analysis carried-out by Horbach (2016): regulations and cost-savings are the main motivations of eco-innovation, while subsidies are relevant for innovations reducing CO2 emissions.

This type of analysis based on micro-data is a relevant way of assessing the effects of supply-side instruments on innovation with environmental benefits. Eco-innovation has a cross-cutting nature. The traditional macro-economic indicators largely based on sectors make it difficult to understand the private R&I performance on environment. The CIS allows overcoming this difficulty. The data show that enterprises supported by the Framework Programme integrate better the environmental aspects in their innovations. Most innovations introduced by enterprises supported by FP6/FP7 have an environmental-friendly component. This allows an interesting conclusion. Even if FP6 and FP7 had not a societal challenge orientation as strong as Horizon 2020, the calls for proposals already focused on green or, at least, resource-efficient technologies. The CIS seems to confirm a positive impact of the Framework Programme from the environmental point of view. It demonstrates the systemic nature of eco-innovation: sustainability principles are largely embedded in the whole FP, and not only in specific themes or societal challenges.

4. ECONOMIC IMPACT OF FP7 PARTICIPATION

The CIS provides also data about the turnover of enterprises and about the percentage of such turnover coming from innovations, especially those new to the market (variable TURNMAR) and those only new to the firm (variable TURNIN).

This information allows to:

i. See whether there are significant differences in terms of turnover between enterprises funded by FP7 and those not funded.
ii. Observe the relative importance of innovation, in terms of turnover, for companies funded by FP7 and for enterprises not supported by the Framework Programme.
iii. Estimate the economic impact, always in terms of turnover, of FP7 funding.

Table 6 shows that there are very significant differences in terms of average turnover between companies funded by FP7 and those not funded. FP7 used to finance enterprises whose turnover is 12 to 14 times bigger. Standard deviations within each category are, of course, huge.

Table 6: Enterprises funded by FP7 vs. enterprises not funded: Differences in terms of average turnover

CIS 2008	Turnover 2006 (€ million)	T-test significance (method)	Turnover 2008 (€ million)	T-test significance (method)
	38.5	<0.0001 (Satterthwaite)	45.7	<0.0001 (Satterthwaite)
	543.8		614	
CIS 2010	**Turnover 2008 (€ million)**	**T-test significance (method)**	**Turnover 2010 (€ million)**	**T-test significance (method)**
	46.05	<0.0001 (Satterthwaite)	43.1	<0.0001 (Satterthwaite)
	659.7		643.6	
CIS 2012	**Turnover 2010 (€ million)**	**T-test significance (method)**	**Turnover 2012 (€ million)**	**T-test significance (method)**
	61.2	<0.0001 (Satterthwaite)	66.6	<0.0001 (Satterthwaite)
	743.1		814.4	

These enormous differences are due to the fact that FP7 tended to support large enterprises (between 34% and 38%, see table 2), while small SMEs represent around half of the companies not supported by FP7 in the survey (48.9% in CIS 2008, 52.5% in CIS 2010 and 48.25% in CIS 2012).

Table 7 shows the impact of innovation activities in terms of percentage of turnover, for both enterprises funded by FP7 and those not funded. Once again, we observe that innovative enterprises supported by the Framework Programme perform better, especially when they introduce new products to the market. Interestingly, the proportion of turnover coming from new to the market innovations increased from the CIS 2008 to the CIS 2012, both for enterprises supported by FP7 and, to a larger extent, for those not funded.

Table 7: Percentage of turnover coming from innovations, funded by FP7 enterprises vs. non-funded

CIS 2008	New to the market innovation (% turnover)	T-test significance (method)	New to the firm innovation (% turnover)	T-test significance (method)
Not funded FP7	5.2	<0.0001 (Satterthwaite)	8.1	0.0004 (Satterthwaite)
Funded FP7	17.4		15.4	
CIS 2010	**New to the market innovation (% turnover)**	**T-test significance (method)**	**New to the firm innovation (% turnover)**	**T-test significance (method)**
Not funded FP7	8.5	<0.0001 (Satterthwaite)	12.2	<0.0001 (Satterthwaite)
Funded FP7	18.7		15.4	
CIS 2012	**New to the market innovation (% turnover)**	**T-test significance (method)**	**New to the firm innovation (% turnover)**	**T-test significance (method)**
Not funded FP7	10.8	<0.0001 (Satterthwaite)	15.7	0.69 (Satterthwaite)
Funded FP7	19.1		15.4	

Instead, the percentage of turnover from new to the firm innovations remained stable for the FP7-funded sub-sample, while other enterprises increased their figure overtime. In the CIS 2012, the differences between the two groups are statistically insignificant.

This trend, to be confirmed in CIS2014, may indicate a positive evolution of European enterprises towards innovation activities.

To estimate the economic impact of innovation activities, the authors have created new variables based on each CIS survey data:

- TURMAREUR: Turnover coming from new or significantly improved products introduced to the market, in euros. It results from TURNMAR * TURN08 (for CIS 2008; TURNMAR * TURN10 for CIS 2010 and TURNMAR * TURN12 for CIS 2012).
- TURNINEUR: Turnover coming from new or significantly improved products to the firm only, in euros, calculated by analogy (e.g. TURNIN * TURN08).
- TOTTURNEUR: Turnover coming from new or significantly improved products to the market or to the firm only, in euros, i.e. TURNMAREUR + TURNINEUR.

Table 8 compares the results of companies funded by FP7 with those not funded. Again, differences between the two categories of enterprises are statistically significant – and huge. However, as shown by the previous tables, such differences are the result of a "size effect" (FP7-funded innovators use to have a much higher turnover, mainly due to their size) and a "turnover effect", as well as the interaction of those variables. This is confirmed by two-ways ANOVA between the variables size of enterprise, FP7 support (FUNRTD) and each of the three variables newly created, in CIS 2008, CIS 2010 and CIS 2012 (see Annex 3). The means suggest that the "size effect" is stronger than the "turnover effect".

Table 8: Average turnover coming from innovations, funded by FP7 enterprises vs. non-funded, in euro

CIS 2008	New to the market innovation (derived turnover, million €)	T-test significance (method)	New to the firm innovation (derived turnover, million €)	T-test significance (method)	Total innovation (derived turnover, million €)	T-test significance (method)
Not funded FP7	3 (N=61,587)	<0.0001 (Satterthwaite)	4.1 (N=61,567)	<0.0001 (Satterthwaite)	7.1 (N=61,567)	<0.0001 (Satterthwaite)
Funded FP7	67.3 (N=1,591)		83.4 (N=1,601)		154.1 (N=1,554)	
CIS 2010	**New to the market innovation (derived turnover, million €)**	**T-test significance (method)**	**New to the firm innovation (derived turnover, million €)**	**T-test significance (method)**	**Total innovation (derived turnover, million €)**	**T-test significance (method)**
Not funded FP7	5.9 (N=36,454)	<0.0001 (Satterthwaite)	6.6 (N=36653)	0.0007 (Satterthwaite)	12.2 (N=36,204)	<0.0001 (Satterthwaite)
Funded FP7	70 (N=1,309)		122 (N=1,323)		196.8 (N=1,284)	
CIS 2012	**New to the market innovation (derived turnover, million €)**	**T-test significance (method)**	**New to the firm innovation (derived turnover, million €)**	**T-test significance (method)**	**Total innovation (derived turnover, million €)**	**T-test significance (method)**
Not funded FP7	6.6 (N=23,035)	<0.0001 (Satterthwaite)	10 (N=23,030)	0.0007 (Satterthwaite)	16.6 (N=22,804)	<0.0001 (Satterthwaite)
Funded FP7	77.8 (N=1,335)		114.6 (N=1,313)		196.3 (N=1,294)	

5. CONCLUSIONS

The European Commission expends large amounts of money in studies aimed at evaluating the Framework Programmes and their impacts. Only the evaluation studies launched for the preparation of the Interim Evaluation of Horizon 2020 sum-up a budget of €1.5 million – and this figure does not include the evaluation actions carried-out at thematic level (e.g. for each Horizon 2020's Societal Challenge, Leadership in Enabling and Industrial Technologies, etc.), which represent at least 14 studies more, conducted mainly through public procurement or expert groups. The ongoing evaluation activities include, for instance, an "Expert Group on evaluation methodologies for the interim and ex-post evaluations of Horizon 2020", with a budget of €0.7 million . With this ambitious study, the Commission seems to implicitly recognise that, despite the significant improvements in recent years, the evaluation and monitoring system of the Framework Programmes still requires new approaches and sources of evidence.

One of the most critical areas for the evaluation of Horizon 2020 remains the impact on innovation. It is particularly relevant in the current political context: innovation is considered a critical driver to create growth and jobs, and there are increasing pressures to guarantee that public investments are accountable and successful. Therefore the ongoing Horizon 2020 could be an opportunity to better collect and measure innovation outputs and outcomes. Few innovation-related indicators are collected directly from projects: only patents, prototypes and testing activities and, for the industrial leadership part of the programme, the share of participating firms introducing innovations new to the company or to the market. This is clearly insufficient and obliges the Commission services to use very complex econometric models, based on several assumptions and disconnected from actual projects' outputs, to estimate the economic impact of the programme (see for instance Martinuzzi et al. 2015, pp.59-60).

This paper demonstrates that the Eurostat's Community Innovation Survey is a very valuable source of information to assess and quantify the impact of the Framework Programme on innovation.

Innovation performance of FP7-funded innovative enterprises and their characteristics

The CIS data show that innovative enterprises supported by FP7 perform better than those not financed by the programme:

- Between 73% and 80% of them introduced new products or services to the market, compared with 32% to 43% of innovative companies not financed by FP7.
- Between 67% and 71% introduced products or services new to the firm only, compared with 42% to 59% of innovative companies not financed by FP7.
- Between 39% and 49% introduced processes new to the market, while the figures are 13% to 20% amongst of innovative companies not financed by FP7.

These data show that FP7-funded innovative enterprises performed best in exploiting new to the market products or services – especially compared with products and services that are just new to the firm.

The CIS data permit also to characterise the FP7-funded enterprises that exploited their innovations on the market:

- Large innovative firms perform better in introducing products and services that are new to the market. Differences are in most cases statistically significant, but not that high (e.g. in CIS 2008, 79% of large innovative enterprises funded by FP7 introduced a new product or service to the market, while the figure for SMEs was around 70%; the gap is lower in successive CIS editions).
- The country where enterprises are based influences more the introduction of new products, services or processes to the market or new to the firm than the size of enterprises. The Czech Republic, Spain, France or Portugal perform best in terms of introduction of new products or services to the market. Participating in FP7 has always a statistically significant positive impact. In Germany, for instance, the gap between FP7-funded and not funded enterprises is very significant, ranging from 23%-27% (non FP-funded) to 66% to 74% (FP funded). In most cases, FP7 participation increases the performance of companies by countries by 25% to 50%.
- The differences in terms of innovation performance between countries are lower when the companies are supported by FP7. This suggests a cohesive role of the Framework Programme for innovative enterprises as a consequence of collaborative R&I activities.
- The Manufacturing sector (NACE C) provides the majority of new to the market product innovations of the European economy, followed by Information and Communication (NACE J, around 10-14%) and by "Wholesale, retail and repair of vehicles" and "Professional, scientific, technical activities" (NACE G and M respectively, slightly less than 10% each). Amongst innovative companies supported by FP7, the sectors Manufacturing, Professional and scientific activities, as well as ICT cover 90% of the new products to the market. The manufacturing sector represents a share of more than 50%, Scientific and technical services around 25% and ICT 12-14%. However ICT enterprises were the most successful in terms of introduction of new products to the market.
- Based on the CIS 2008 only, innovative companies supported by FPs delivered more environmental-friendly products and services. They introduced almost three times more eco-innovations (+183.5%) than enterprises that did not benefit from FP's support.

The CIS shows also that participation in FP7 has a positive economic impact measured in terms of turnover. Innovative firms supported by FP7 present a proportion of sales of new to the market products twice or three times higher than companies not funded by the Framework Programme. The paper calculates what this represents in euros, but the amounts are heavily influenced by the size of (some) enterprises that received FP7 support.

27 Horizon 2020 – Work Programme 2016-2017, section 19 (Dissemination, exploitation and evaluation). European Commission Decision C(2016)4614 of 25 July 2016, at: http://ec.europa.eu/research/participants/data/ref/h2020/wp/2016_2017/main/h2020-wp1617-comm-diss_en.pdf

CIS issues to assess the innovation impact of the Framework Programmes

The CIS also presents caveats that limit its capacity to evaluate the evaluation impact of the Framework Programme: geographical coverage of publicly available data, years of reference that do not exactly coincide with the FPs' timeframe, etc. The analysis of the data has shown that main issues are linked with and limited by the questionnaire design and confidentiality rules:

- The question that allows identifying enterprises supported by FP7 comes from a previous filter. Enterprises that are not involved in innovation activities do not answer such question. This means that the FUNRTD variable, key to assess FP7, covers "FP7 innovators" only, i.e. enterprises that carry-out innovation activities and received an FP7 grant. We cannot assume that all private for profit organisations supported by the Framework Programme are involved in innovation and FUNRTD cannot be considered a priori a sub-sample of FP7 participants. This makes extrapolating results to the overall FP impossible.
- Indeed, also due to the position of the question on FP7 support, marketing and organisational innovations are excluded (OECD 2015). "FP7-innovators" cover product, service and/or process innovations only, which does not cover the whole spectrum of possible innovations.
- The confidentiality rules imply that, when we cross more than two variables (e.g. FUNRTD and sector, or country), information for some categories is lost. The results are then incomplete and, indeed, more detailed data, for example by NACE 2 or 3 digits become unavailable.

Of course, one major issue is that correlations do not mean causality. The CIS data demonstrate that FP7-funded innovative enterprises perform better, but this does not mean that it is thanks to FP7. It can be argued that FP7 attracted the most innovative enterprises, which also benefit from international R&I cooperation financed by the EU, amongst other factors. This is however a problem that can hardly be solved.

Despite its caveats, the CIS is a gold mine for R&I policy and should be further exploited by the Commission services. It provides quantitative data on impacts of R&I on innovation, measured in terms of exploitation: products, services or processes new to the market and/or to the firm, as well as the turnover obtained caused by the introduction of new products, services and processes. Even if the data are incomplete or not detailed enough (for example, to analyse the specific impacts by FP7-Cooperation Theme or Horizon 2020-Societal Challenge), important information is provided.

In addition, the CIS data are accessible for free by research entities and are easy to use with a basic statistical knowledge. The analysis that is presented in this paper is indeed based on simple cross-tables (contingency tables, t-tests and ANOVAs), trying to respect the parsimony principle and making the results as easy to understand as possible.

6. REFERENCES

LITERATURE

Arundel, Anthony (2007) 'Innovation survey indicators: what impact on innovation policy?". In: OECD (ed.) Science, Technology and Innovation Indicators in a Changing World. Responding to Policy Needs. Paris: OECD.

Connolly, Niam et al. M.(2014) Ex-Post Evaluation of FP7, Cooperation Programme Theme: "Environment (including Climate Change). Brussels: European Commission. At: http://bookshop.europa.eu/en/ex-post-evaluation-of-fp7-cooperation-programme-theme-environment-including-climate-change--pbKI0614224/

Eco-Innovation Observatory (EIO, 2013) A systemic perspective on eco-innovation. Brussels: European Commission/EIO Thematic Report.

European Commission (2015) From Niche to Norm. Suggestions by the Group of Experts on a 'Systemic Approach to Eco-Innovation to achieve a low-carbon, Circular Economy'. Brussels: European Commission.

Eurostat (2008) NACE Rev. 2. Statistical classification of economic activities in the European Community. Luxembourg: OPOCE.

Fisher, Robbert; Polt, Wofgang and Vonortas, Nicholas (2009) The impact of publicly funded research on innovation: An analysis of European Framework Programmes for Research and Development. Luxembourg: OPOCE.

Hassan, Emmanuel et al. (2013) Testing Horizon 2020 performance indicator on SMEs involved in innovation projects under FP7. Brussels: Science-Metrix, produced for the European Commission, DG Research and Innovation.

Horbach, Jens (2016) 'Empirical determinants of eco-innovation in European countries using the Community Innovation Survey'. Environmental Innovation and Societal Transitions, Vol. 19, pp.1-14.

Joint Research Centre (various editions) EU Industrial Scoreboard reports. Brussels: European Commission, Joint Research Centre. At: http://iri.jrc.ec.europa.eu/scoreboard.html

Martinuzzi, André et al. (2015)Commitment and Coherence, essential ingredients for success in science and innovation. Ex Post Evaluation of the 7th Framework Programme (2007-2013). Brussels: European Commission. At: http://ec.europa.eu/research/evaluations/pdf/fp7_final_evaluation_expert_group_report.pdf#view=fit&pagemode=none

Mazzanti, M. et al. (2016) 'Firm surveys relating Environmental Policies, Environmental Performance and Innovation. Design challenges and insights from empirical application', OECD Environment Working Papers No. 103. Paris: OECD. At: http://www.oecd-ilibrary.org/environment/firm-surveys-relating-environmental-policies-environmental-performance-and-innovation_5jm0v405l97l-en?crawler=true

Muldur, Ugur. et al. (2006) A New Deal for and Effective European Research Policy. The design and impacts of the 7th Framework Programme. Dordrecht: Springer.

OECD (2015) Oslo Manual. Guidelines for collecting and interpreting innovation data. Paris: OECD.

OFFICIAL DOCUMENTS

Decision 1982/2006 of the European Parliament and the Council of 18 December 2006 concerning the Seventh Framework Programme of the European Community for research, technological development and demonstration activities (2007-2013). At: http://cordis.europa.eu/documents/documentlibrary/90798681EN6.pdf

Communication from the European Commission (2008) A European Economic Recovery Plan, COM (2008)800 final. At: http://ec.europa.eu/economy_finance/publications/publication13504_en.pdf

Communication from the European Commission (2010) Europe 2020: A strategy for smart, sustainable and inclusive growth, COM (2010)2020 final. At: http://ec.europa.eu/eu2020/pdf/COMPLET%20EN%20BARROSO%20%20%20007%20-%20Europe%202020%20-%20EN%20version.pdf

Communication from the European Commission (2010) Europe 2020 Flagship Initiative Innovation Union, COM (2010) 546 final. At: http://bookshop.europa.eu/en/ex-post-evaluation-of-fp7-cooperation-programme-theme-environment-including-climate-change--pbKI0614224/

Commission Staff Working Document accompanying the Communication from the Commission 'Horizon 2020 - The Framework Programme for Research and Innovation', SEC(2011)1487 final. At: http://ec.europa.eu/research/horizon2020/pdf/proposals/horizon_2020_impact_assessment_report.pdf

Regulation 1291/2013 of 11 December 2013 of the European Parliament and the Council establishing Horizon 2020 - the Framework Programme for Research and Innovation (2014-2020). At: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2013:347:0104:0173:EN:PDF

Council Decision of 3 December 2013 establishing the specific programme implementing Horizon 2020 - the Framework Programme for Research and Innovation (2014-2020). At:
Communication from the European Commission (2015) Closing the loop – An European action plan for the Circular Economy, COM(2015) 614 final. At: http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52015DC0614

Communication from the European Commission (2016) Communication on the Response to the Report of the High Level Expert Group on the Ex Post Evaluation of the Seventh Framework Programme, COM (2016)5 final. At: https://ec.europa.eu/research/evaluations/pdf/archive/fp7-ex-post_evaluation/commission_communication_1_en_act_part1.pdf

Horizon 2020 – Work Programme 2016-2017, section 19 (Dissemination, exploitation and evaluation). European Commission Decision C(2016)4614 of 25 July 2016, at: http://ec.europa.eu/research/participants/data/ref/h2020/wp/2016_2017/main/h2020-wp1617-comm-diss_en.pdf

AUTHORS

corresponding author:
JESÚS ALQUÉZAR SABADIE
European Commission - DG Research and Innovation
CDMA 3/175
1049 – Brussels (Belgium)
e-mail: jesus-maria.alquezar-sabadie@ec.europa.eu

CLAIRE KWIATKOWSKI
European Commission - DG Research and Innovation

ACKNOWLEDGEMENT/DISCLAIMER:

This paper is based on data from Eurostat's Community Innovation Survey, editions 2008, 2010 and 2012. It reflects the views only of the authors; neither the European Commission nor any person acting on behalf of the Commission is responsible for the use which might be made of the following information. The responsibility for all conclusions drawn from the data lies entirely with the authors.
The authors would like to thank Geneviève Villette (Eurostat) for her valuable support, as well as the different people who read, reviewed and commented the paper.

STATISTICAL ANNEX

Please find the statistical annex for this paper on the following website: http://repository.fteval.at/id/eprint/305

ASSESSING THE IMPACT OF PUBLIC FUNDING AND TAX INCENTIVES IN RUSSIA: RECIPIENT ANALYSIS AND ADDITIONALITY EFFECTS EVALUATION[1]

YURI SIMACHEV , MIKHAIL KUZYK AND NIKOLAY ZUDIN

ABSTRACT

So far a considerable number of studies have used the concept of additionality as basic evaluation approach but none of them paid attention to additionality effects of the Russian innovation policy yet. In this study we performed a microeconomic evaluation of the industrial firms' public innovation support in Russia focusing on its two key toolbox elements: direct funding and tax incentives. Based on the data from a questionnaire survey of top executives of Russian manufacturing firms from 2015 we identified and evaluated the profiles and the performance of recipients of direct funding and tax incentives. We also assessed the "relative" additionality - the additionality of a concrete instrument for a particular firm relative to all other used instruments - with propensity score matching. The results show that generally Russian industrial innovation policy tends to target sufficiently large and long-operating companies. In terms of effects we have confirmed not only the importance of the fiscal support in providing main aspects of additionality but also its significance in crowding-out private investment. Our results suggest tax incentives do hardly contribute to additionality of any kind which is especially unusual regarding input additionality. One should also point out a relatively small impact of public support on science-business cooperation which is quite unexpected in view of the substantial effort provided by the Russian government in enhancing its development.

INTRODUCTION

In recent years the attention to innovation policy evaluation, particularly as a means of learning and also as a search for best practice, has intensified in many countries. What has been successful in one country may be counterproductive in the other, so the problem identification should be combined with certain „experiments" over solution methods combined with learning processes development (Rodrik 2008; Chaminade et al. 2009). By now an extensive experience in assessing the impact of public policies on fostering companies' innovation activities has been accumulated. There is also significant progress in the development of methodologies concerning the assessment of science-technology and innovation policies (Crespi et al. 2011). Objectives of the innovation policy evaluation have become more complex due to the need of recording a substantial amount of different factors and effects, including non-economic ones. The evolution of evaluation techniques develops in several directions (OECD 2012):

(1) establishment of frameworks and conditions, formation of the evaluation culture, and - sometimes - the development of assessment legislation;
(2) expansion of assessing institutions and their coordination, distribution and improvement of assessment practices;
(3) formation of the base and infrastructure for evaluation - determination of standards and methodologies, combining assessment with Key Performance Indicators (KPI), accumulation of policy implementation data, support for the evaluation expert community.

The following features of a modern evaluation practice of innovative instruments can be distinguished:

(1) regular assessment, cross-country comparison of results;
(2) long observation periods (over 10 years), maintenance of extensive detailed statistical databases used for performance evaluations; openness of assessment procedures to capture new effects.
(3) the complexity and ambiguity of estimates (e.g. considering substitution effect), presence of significant time lags over output effects (4-6 years), high heterogeneity of the impact of incentive mechanisms; substantial econometric problems; preparation and submission of guidelines on principles and problems of evaluation;
(4) openness, publicity of assessment results; practical use of the assessment results for decision-making at government level – the spread of best practices; drawing lessons: in particular, making decisions to stop, clarify or expand various programs, mechanisms and instruments aimed at fostering innovation activity

It must be admitted that the Russian public innovation policy evaluation system in comparison with the best foreign examples turns out to be quite imperfect and unbalanced – due to the excessive emphasis on direct results of support and lack of attention to the process of drawing lessons.

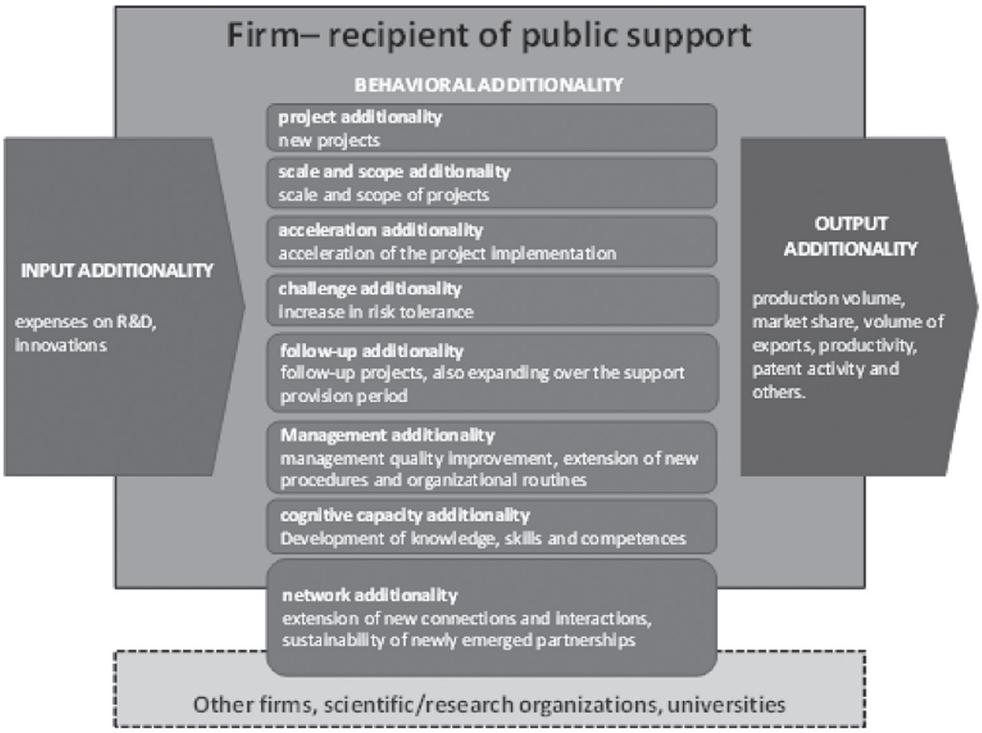
1 The paper was presented at the International RTI Policy Evaluation Conference "Open Evaluation 2016", 24th November 2016

CONCEPT OF ADDITIONALITY: MAIN POINTS

In recent years the concept of additionality has become the basic evaluation approach of the innovation policy toolbox. In the context of government intervention the notion of additionality involves a comparison of the real situation of receiving government support with a hypothetical scenario of what would have happened if no support had been provided.

The central element of the additionality concept is the change in specific indicators and company characteristics achieved thanks to government support. For this reason additionality is often classified by the type of considered indicators and thus is divided into input, output and behavioral additionality. It must be added that in economic literature a significant portion of innovative development problems lies in the features of economic agents innovative behavior: lack of responsiveness to new knowledge, low level of cooperative activities etc. (Gok, Edler 2011). Thus, behavioral additionality should be considered more closely. Also more detailed subclasses of each type of additionality exist, which are briefly represented in figure 1. To save time we will not dig into them but instead would highlight some important points concerning the basic concept of additionality.

Figure 1 – Main types of additionality



Sources: based on Buisseret et al. (1995); Georghiou (1997, 2002); Bach, Matt (2002); Georghiou et al. (2002); Rye (2002); Falk (2004); Georghiou, Clarysse (2006); Idea Consult (2006); Hsu et al. (2009); Gok (2010); Roper, Hewitt-Dundas (2012); Viljamaa et al. (2013); Wanzenbock et al. (2013); Lohmann (2014); Neicu et al. (2014)

In general, along with the main and obvious advantage of the additionality concept which lies in operating with "clean" results of public support not being achieved in the absence of the latter, one can identify a number of other important arguments in favor of this approach:

- a wide range of considered effects, including hardly formalized "quality" results of support, such as development of partnerships and competences;
- consideration of not only the direct influence of public support on a recipient, but also of the indirect impact on his/her partners in science-industrial cooperation;
- consideration of government support effects not only in the period of its provision but also after its termination is essential, firstly, to test the robustness of the results and, secondly, due to the fact that these effects often occur with a significant time lag, sometimes even a few years after provision of government support (Lopez-Acevedo, Tan 2010; Crespi et al. 2011a).

Thus, the use of the additionality concept as the basis for the analysis of public support effectiveness enables to minimize the risks of overvaluation (due to the orientation on a net effect) and of undervaluation (what is even more important from the standpoint of identifying and disseminating best practices).

CONTEMPORARY EMPIRICAL EVIDENCE ON PUBLIC SUPPORT ADDITIONALITY: BRIEF OVERVIEW

So far a considerable number of studies assessing the effects of innovation policy on firms' activity with the use of the concept of additionality has taken place. However, none of the known studies paid attention to additionality of the Russian innovation policy. In most of them the objects of analysis were various instruments of government funding solely or (more rarely) financial support for innovation in any form as the direction of the public innovation policy. Making no claim to cover all the existent empirical evidence on additionality, we, nevertheless, can identify some common features and patterns regarding the additionality of financial support on the basis of the findings of about thirty studies.

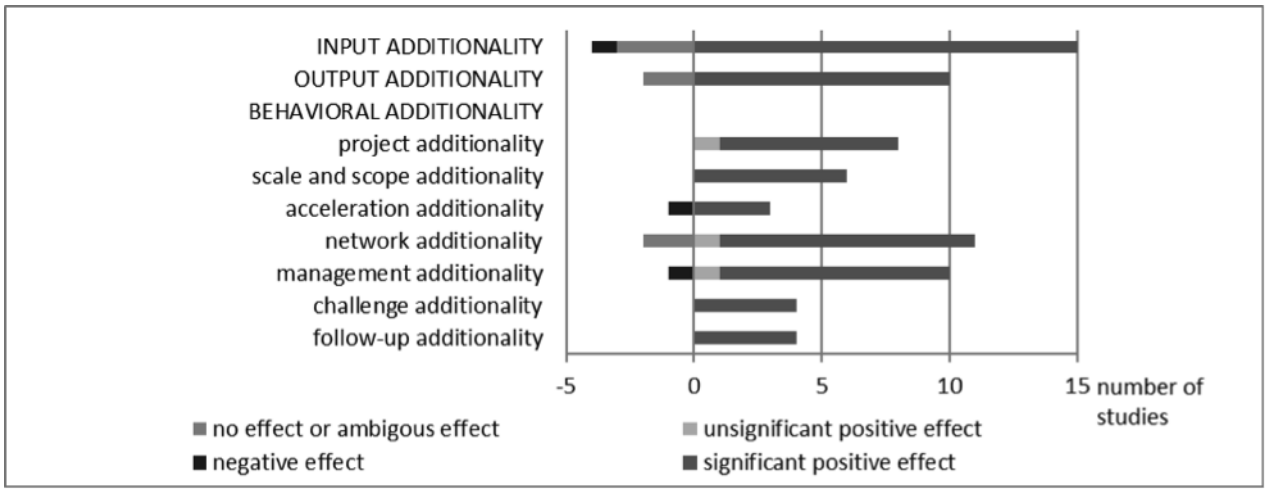
Input and output effects of public support as well as network additionality became most often subjects of analysis (figure 2). Three other types of behavioral additionality – acceleration, follow-up and challenge - in contrast, have relatively rarely attracted the attention of researchers. Finally, despite the fact that cognitive capacity additionality is often considered in theoretical works, in practice it is usually either not included in the scope of the empirical analysis or considered as a part of the follow-up or management additionality.

In most cases the results obtained by the researchers confirmed the presence of different additionality effects. However, there are important single reverse examples.

For instance, the study of Marzucchi, Montresor (2013) dedicated to the analysis of the results of financial support provided at regional and national levels for companies' innovation activities in manufacturing sector of two European countries - Italy and Spain - revealed in both countries the absence of input additionality for regional innovation funding in contrast to its presence at the national level. With regard to Italy the authors discovered significant positive impact of government support (both at regional and national levels) on process innovations, but negative impact on product innovation financing at the regional level. Researchers believe that obtained results can be explained by the fact that regional support stimulates more the "deepening" of the innovation activity rather than the production of new products. In addition, a negative relationship between regional funding and improvement of employees' qualification could be identified in case of the Italian companies. Also the same consistent pattern was found for networking with other companies aimed at obtaining information, while in the case of information networking with the scientific organizations the effect of regional funding has been, in contrast, positive.

In the study of Montmartin, Herrera (2015) devoted to the analysis of the public financial support and tax incentives on the scale of private funding of R&D on the basis of data for 25 OECD countries over a twenty-year period (1990-2009), the authors came to the conclusion that tax incentives at the country level increase the intensity of business expenditure on R&D, while direct government funding leads rather to the opposite result. In a more „localized" study (Montmartin et al. 2015) on the basis of data from 94 regions of mainland France for 2001-2011, the authors analyzed the direct and indirect impact of financial support provided at the regional, national and supranational (EU) level on private R&D spending. The results suggest that a significant input additionality

Figure 2 –Additionality of direct financial support for innovation activities of companies - generalization of empirical evidence



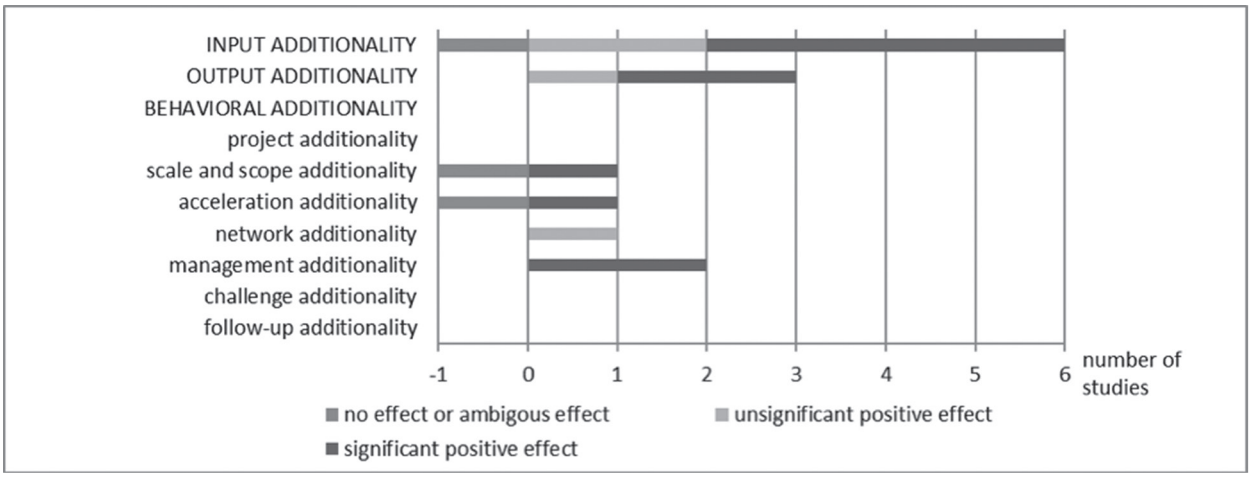
Sources: based on Ðallejón, Quevedo (2005), Pegler (2005), Georghiou et al. (2005), OECD (2006), González et al. (2005), Czarnitzki, Licht (2006), Fier et al. (2006), Falk (2007), Busom. Ribas (2008), Hsu et al. (2009), Gelabert et al. (2009), Clausen (2009), Knockaert, Spithoven (2009), Idea Consult (2009), Baghana (2010), Wanzenbock et al. (2011), Catozzella, Vivarelli (2011), Marzucci, Montresor (2012), Antonioli et al. (2012), Lucena, Afcha (2013), Lohmann (2014), Hud, Hussinger (2014), Montmartin, Herrera (2015), Cantner, Kösters (2015), Montmartin et al. (2015).

was revealed only for the national government subsidies, while the corresponding additionality was statistically insignificant for EU financing and regional subsidies.

Quite interesting and unexpected results have been received by Lohmann (2014) through in-depth interviews with project managers in the airline industry: government participation in financing of projects, contrary to expectations, did not lead to the reduction in terms of their implementation time, but on the contrary - to prolongation, meaning that the acceleration additionality is negative. This is determined by the long period of provision of subsidies which significantly exceeded the typical duration of supported projects. At the same time, however, a number of respondents described the increase in duration as a result of public support for projects as a positive effect, which allowed carrying out more detailed projects and thus increasing the chances of their successful implementation.

Tax incentives are considerably less the object of additionality studies. As a rule, researchers only considered the input additionality and obtained an empirical evidence of its presence (Figure 3). In the above mentioned research (Montmartin et al. 2015), the authors came to the conclusion that the direct positive effect of tax incentives for a particular region approximately balances the indirect negative effect for the other regions.

Figure 3 – Additionality of tax incentives for innovation activities of companies - generalization of empirical evidence



Sources: based on Poot et al. (2003), Hægeland, Møen (2007), Catozzella, Vivarelli (2011), Lokshin, Mohnen (2012), Zoran, BotriÐ (2013), Neicu et al. (2014), Montmartin, Herrera (2015), Freitas et al. (2015), Montmartin et al. (2015).

OBJECTIVES, METHOD AND DATA

The aim of our study is to perform a microeconomic evaluation of support provided to industrial firms in Russia by focusing on two main instruments: direct funding and tax incentives. The usage of these two instruments for the analysis is quite straightforward as they are traditionally viewed as key elements of the national innovation policy toolbox (e.g. David et al. (2000); OECD (2015)) and are well ahead of other instruments in terms of "coverage" and number of firms supported (Kuzyk, Simachev 2013).

The objectives of this study were as follows:

- firstly, to identify the "typical profile" of the firms-beneficiaries of the government support policy as a whole and of direct funding and tax incentives in particular;
- secondly, to consider basic input, output and behavioral additionality effects;
- thirdly, to analyze the "relative" additionality of direct financial support and tax incentives.
- Data were collected from a questionnaire-based survey, which addressed top executives of Russian manufacturing firms in September-October 2015.

The organizer of the survey was the Interdepartmental Analytical Center. The field operation was carried out by the publishing and information center „Statistics of Russia". While constructing the sample we targeted public support recipients to meet the research objectives. Thereby we tried to capture in our sample, on the one hand, mostly high technology industries which are relatively more often supported by the government in Russia than low-tech and medium-tech industries (Zudin 2015) and, on the other, large companies because as results of several studies prove they become more often beneficiaries of government support more in comparison with SMEs (Fier, Heneric 2005; Aschhoff 2010; Simachev et al. 2014a). As a result, the final sample consisted of 658 firms, ¾ of which belong to high-tech industries – first of all from the chemical and machine-building complex. The sample is characterized by an equal share of small companies and relatively large enterprises (while the general population is dominated by small firms) (Table. 1). Additionally, it should be noted that in the sample - as in the whole of Russian industry - companies with private ownership prevail, but at the

same time public companies are also widely represented, what is very important from the standpoint of public support distribution analysis and its results. Finally, a major part of the surveyed firms are in a relatively healthy financial condition, but companies experiencing financial problems are also significantly represented in the sample.

Table 1. Sample structure

Characteristic		Share in the sample
Industry	Manufacturing textiles, clothing and footwear	7,45%
	Wood processing, manufacturing of wood products, pulp, paper and paperboard	5,32%
	Chemical production (excluding pharmaceutics)	6,23%
	Manufacturing of pharmaceutical products	4,71%
	Metallurgy, manufacturing of finished metal products	9,73%
	Manufacturing of machinery and equipment (except for machine-tools)	18,84%
	Manufacturing of machine-tools	3,95%
	Manufacturing of electrical machinery and electrical equipment	8,36%
	Manufacturing of computer technology, equipment for processing information, radio, TV and telecommunication	9,42%
	Manufacturing of medical equipment	4,86%
	Manufacturing of control and measuring devices	3,65%
	Automobile production	4,56%
	Shipbuilding	4,10%
	Manufacturing of railway rolling stock	4,86%
	Manufacturing of aircraft	3,95%
Operation period	less than 5 years	8,81%
	5-10 years	16,26%
	10-20 years	26,90%
	more than 20 years	48,02%
Ownership	state and municipal (including the ownership of state-owned corporations)	9,27%
	mixed	5,78%
	private	84,95%
Number of employees	less than 100 emp.	24,77%
	101-200 emp..	22,95%
	201-500 emp..	24,32%
	more than 500 emp..	27,96%
Financial condition	poor	17,93%
	satisfactory	69,91%
	good	12,16%

Sources: prepared by the authors

To determine the composition of the companies that are "consumers" of government support we used a questionnaire in which respondents were asked to mark if they had received budget financing, used tax incentives or other public support instruments and measures in 2013-2015. Actually, the analysis of the public support results and additionality is based on the responses from CEOs for covering a wide range of corresponding effects related to "classical" types of additionality with the exception of management and follow-up additionality (Table 2). It is important to note that one of the main restrictions of the survey was that we only asked one person in a company. CEOs were chosen as they are actually better informed compared to any other single specialist regarding the firm's profile, its position on the market and most importantly the impact of public support of different kinds on its activities and organizational routines.

Table 2. Comparison of the public support effects and different types of additionality

Effect	Additionality
volume of company's investment in new equipment based on its own or borrowed funds has increased	input
volume of company's spending on innovation based on its own or borrowed funds has increased	
volume of company's spending on R&D based on its own or borrowed funds has increased	
company's revenue has increased	output
company's market share on the domestic market has increased	
company's market share on the external market has increased	
production volume of new (improved) products has increased	
profitability of core company's activities has improved	
company's general competitiveness has increased	
import dependence of the company has been reduced	import substitution
a promising new project (projects) was launched	project
public support has allowed to implement a larger project (projects)	scale and scope
public support has allowed to implement a project (projects) with a longer payback period	
public support accelerated the implementation of the project	acceleration
public support reduced risks of project implementation	challenge
development (strengthening) of the company's linkages within scientific and industrial cooperation has occurred	network
public support has allowed to redistribute part of company's funds towards the other areas not related to the subject of support	—

Sources: prepared by the authors

We associate the reduction of risks of innovation activities as a result of public support with the 'challenge additionality' since this effect contributes to the risk "tolerance" of companies and thus to the initiation and implementation of more risky projects. Reducing import dependence of companies (an item which was included in the scope of the analysis due to the active implementation of the import substitution policy in the Russian industry) does not directly correspond to any of the "classic" types of additionality and for this reason we regard it as a separate category. Finally, the redistribution of existing funds towards other areas not related to the subject of public support is nothing more than the well-known and frequently observed "crowding out" effect in national and foreign studies, which is the opposite to input additionality (David et al. 2000; Lach 2002; Chudnovsky et al. 2006; Benavente et al. 2007; Simachev et al. 2015).

We use frequency and regression analysis to identify the specifics of the beneficiaries of public support and reveal its main additionality effects. For a more precise definition of "relative" additionality effects of direct funding and tax incentives we used a propensity score matching (PSM) which is one of the main techniques of analyzing additionality at firm level (e.g. Fier et al. (2006); Baghana (2010); Marzucchi, Montresor (2013); Cantner, Kösters (2015)). An important distinguishing feature of our approach is that we analyzed the additionality of a concrete instrument for a particular firm relative to all other instruments used and therefore could consider the "relative" additionality. This enables us to highlight inherent additionality effects precisely to tax and financial instruments distinguishing them from the "background" of all other elements of the innovation policy toolbox.

It is necessary to pay attention to two important limitations of our study. Firstly, we analyzed the "generalized" streams of public support (direct financial support and tax incentives). More specifically, in our research we did not focus on particular instruments but considered financial support and tax incentives in any form. It seems to us justifiable because these two policies are fundamentally different and companies tend to choose a set of instruments of one kind - financial or tax (Ivanov et al. 2012). Note also that such a "generalized" approach is widely used in modern empirical studies (e.g. Gelabert et al. (2009); Lokshin, Mohnen (2012); Marzucci, Montresor (2012); Hud, Hussinger (2014); Bodas Freitas et al. (2015)).

Secondly, we do not distinguish between federal and regional support. Meanwhile, as noted above, the effects of such support can vary considerably (Marzucchi, Montresor 2013). However, such differences most clearly manifest themselves in the case of financial support. In Russia the financing of innovation activities from the federal budget significantly exceeds the volume of the relevant funding at the regional and local levels (Gorodnikova et al. 2016).

RESULTS

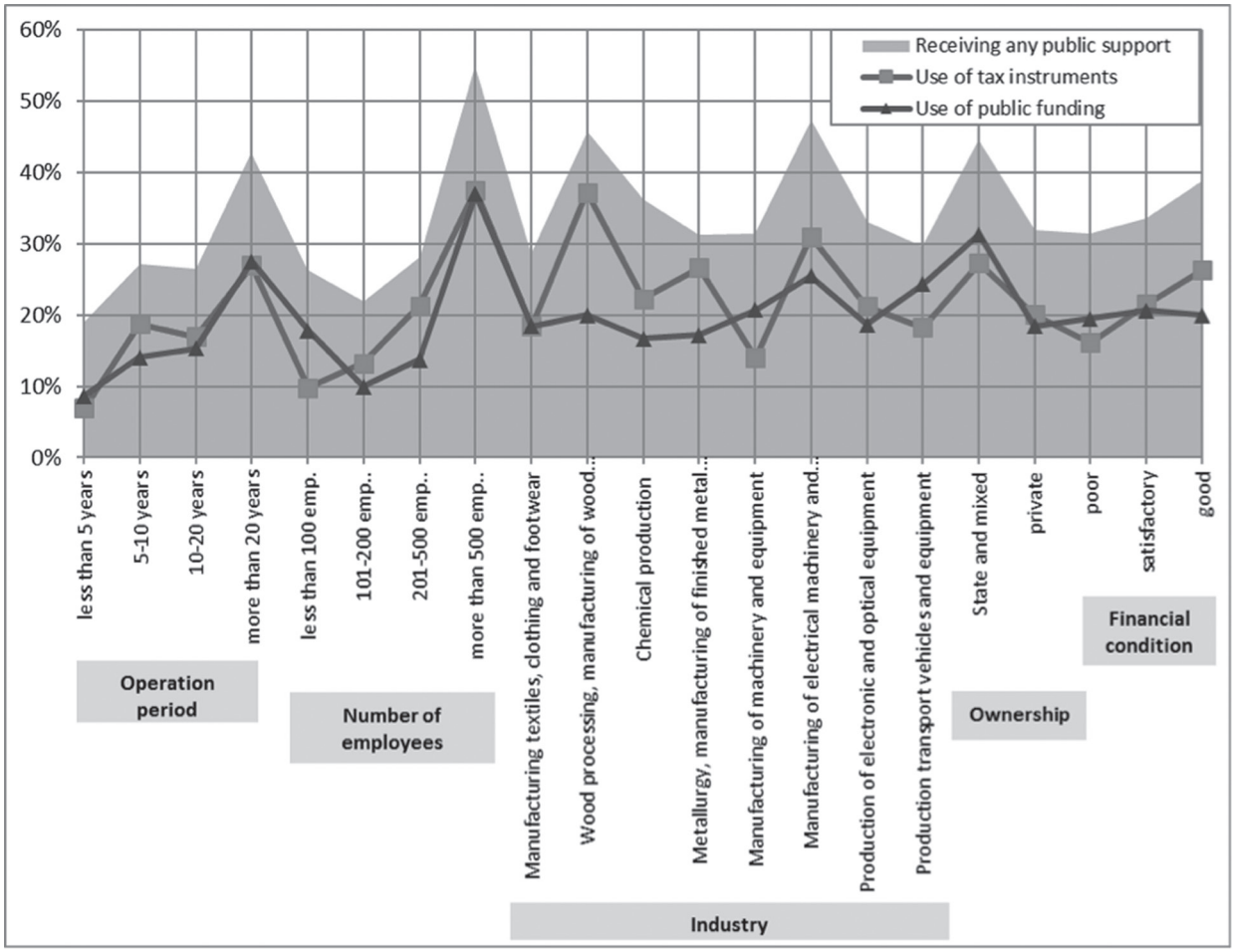
BENEFICIARIES OF PUBLIC SUPPORT - THE TYPICAL PROFILE

Before analyzing additionality results caused by tax and financial instruments we briefly turn to the formation of the typical profile of the beneficiaries of public support. On the whole one third of the sampled companies used some public support instruments in 2013-15; while 20 per cent of companies have received government funding, about the

same – 21 per cent have made use of different tax incentives and 10 per cent of companies received both streams of support.

Large and long operating (over 20 years) companies have more frequently become the recipients of public support than relatively newly established and small (100-200 employees) companies. Note that such a "failure" of the public support in relation to small firms was observed by us earlier (Simachev et al. 2014a).

Figure 4 - Receiving state support in 2013-2015 by companies of different categories - percentage of the total company sample by category



Sources: prepared by the authors

When considering different industries one can note a certain "concentration" of both public support dimensions in the field of electrical equipment production and also a relatively frequent use of tax incentives by wood processing companies in a line with pulp and paper industry companies. Finally, companies with government participation in capital more frequently become the beneficiaries of financial support and tax incentives.

Regression analysis by including the explanatory variables presented in Figure 4 allows us to define more clearly the profile of beneficiaries using different tax and financial support instruments (Table 3.) The use of tax incentives in the review period was most typical for large companies and unusual for SMEs (not more than 200 employees). Large and long-operating companies had the greatest chance to receive government financial support; at the same time, however, small firms also often became the recipients of budget funds.

Table 3. State support for companies in 2013-15 - results of the binary logistic regression model

Independent variables (dummy)		Dependent variables (dummy)		
		Receiving any public support	Use of tax instruments	Use of public funding
Industry	Manufacturing textiles, clothing and footwear	control		
	Wood processing, manufacturing of wood products, pulp, paper and paperboard	+*	+**	
	Chemical production			
	Metallurgy, manufacturing of finished metal products			
	Manufacturing of machinery and equipment			
	Manufacturing of electrical machinery and electrical equipment	+*		
	Production of electronic and optical equipment			
	Production transport vehicles and equipment			
Operation period	less than 5 years			
	5-10 years			
	10-20 years	control		
	more than 20 years	+**		+**
Ownership	State and mixed			
	private	control		
Number of employees	less than 100 emp.		-**	+*
	101-200 emp..		-*	
	201-500 emp..	control		
	more than 500 emp..	+***	+**	+***
Financial condition	poor			
	satisfactory	control		
	good			
Chi-square		78,39***	66,98***	65,44***
N		658		

Notes:
Maximum VIF value – 3,20.
* – significant a 10% level;
** – significant a 5% level;
*** – significant a 1% level.
Sources: prepared by the authors

Interestingly, the regression analysis revealed no statistically significant relationship between private industrial companies and fully public companies or companies with mixed ownership, which is consistent with previous empirical results (Simachev et al. 2014a).

MAIN EFFECTS OF PUBLIC SUPPORT

Talking about input additionality the most widely observed effect is the increase in investment in new equipment based on own or borrowed funds. Interestingly, almost as often respondents mentioned the crowding out effect (see Table 4.).

The most common output effects of public support are the increase of new and improved products produced and increases in revenue, as well as growth of profitability and general company competitiveness. In contrast, least likely public support led to increasing market shares of companies on the external markets. Although not significantly, public policies of any kind also contributed to a decreasing import dependence

of the scrutinized companies. Among the key behavioral effects of public support one can find accelerated project implementation and encouragement of launching new projects. Government support relatively rarely led to the development of scientific-industrial cooperation, which seems very surprising in view of the impressive scale of the public policy in recent years aimed at encouraging network between science and industry (Simachev, Kuzyk 2015).

Input, output, as well as behavioral additionality was effected to a significantly greater extent by financial instruments than by tax incentives (Table 4). Financial support significantly stood out from the other instruments in terms of effects such as growth of investment in new equipment, increase in production volumes of new and improved products, acceleration of project implementation and risk tolerance At the same time, the crowding-out of own private funds through government funds was higher in case of direct financial support.

Table 4. Major results of public support – frequency statistics

	All the recipients of government support	Companies using tax incentives	Companies enjoying public funding
INPUT ADDITIONALITY	35,6%	36,7%	42,5%*
- volume of company's investment in new equipment based on own or borrowed funds has increased	28,4%	25,9%	34,3%**
- volume of company's spending on innovation based on own or borrowed funds has increased	15,3%	17,3%	17,2%
- volume of company's spending on R&D based on own or borrowed funds has increased	15,8%	17,3%	18,7%
OUTPUT ADDITIONALITY	47,3%	48,2%	56,0%***
- company's revenue has increased	18,9%	16,5%	20,9%
- company's market share on the domestic market has increased	13,5%	10,1%*	16,4%
- company's market share on the external market has increased	2,7%	2,9%	2,2%
- production volume of new (improved) products has increased	21,2%	19,4%	27,6%***
- profitability of core company's activities has improved	18,0%	18,0%	17,9%
- company's general competitiveness has increased	18,9%	16,5%	22,4%
IMPORT SUBSTITUTION ADDITIONALITY - import dependence of the company has been reduced	4,1%	4,3%	4,5%
BEHAVIORAL ADDITIONALITY	44,6%	46,0%	55,2%***
- a promising new project (projects) was launched	19,8%	21,6%	24,6%**
- public support has allowed to implement a larger project (projects)	11,7%	12,9%	14,2%
- public support has allowed to implement (projects) with a longer payback period	12,6%	15,1%	14,2%
- public support accelerated implementation of the project	21,6%	20,9%	26,9%**
- public support enabled to reduce risks of project implementation	13,5%	15,1%	17,2%*
- development (strengthening) of the company's linkages within the scientific and industrial cooperation has occurred	8,1%	8,6%	9,7%
CROWDING-OUT EFFECT - public support has allowed to redistribute part of company's funds towards other areas not related to the subject of support	29,3%	27,3%	37,3%***
N	222	139	134

Notes:

The significance of differences, the Chi-square test
* – significant a 10% level;
** – significant a 5% level;
*** – significant a 1% level.
Sources: prepared by the authors

RELATIVE ADDITIONALITY OF TAX INCENTIVES AND PUBLIC FUNDING

In this section we would address the "relative" additionality, i.e. the additionality of a concrete instrument for a particular firm relative to all other used instruments, to capture "net" additionality effects.

For a more precise identification and comparison of the effects generated by tax respectively financial support, we use the following typical algorithm[2]:

1. The first step involved the assessment of the two sets of expected probabilities (propensity scores) of using tax incentives or of obtaining financial support by building bivariate logistic regressions with a "standard" set of control variables for the sampled companies (see Table 3). Since the questionnaire about public support effects was addressed only to recipients evaluation was carried out under the sub-sample of companies which obtained government support of any form in 2012-15.
2. Secondly, for the analyzed two dimensions of public policy (tax incentives and direct financial support), pairs of most similar recipients and non-recipients were identified. Pairs were formed by nearest neighbor matching on the basis of propensity score variables created at the first step.
3. Thirdly, the average effects on the treated companies (or ATTs) for tax and financial support were estimated (for more details see Newey (2009)). It is important to note that initially this effect is described by the following functional dependence

$$ATT_{ij} = E(Y_{ij}^1 - Y_{ij}^0 \| S_i = 1) = E(Y_{ij}^1 \| S_i = 1) - E(Y_{ij}^0 \| S_i = 1) \quad (1)$$

где i – considered dimension of support (tax or financial); j – specific effect of the support; Y_{ij}^1 - indicator of presence or absence of the effect in case of receiving support of j type i; Y_{ij}^0 - the corresponding indicator in the hypothetical situation, if this type of support has not been received; Si – indicator of obtaining support of a specific type i (1 – presence, 0 – absence).

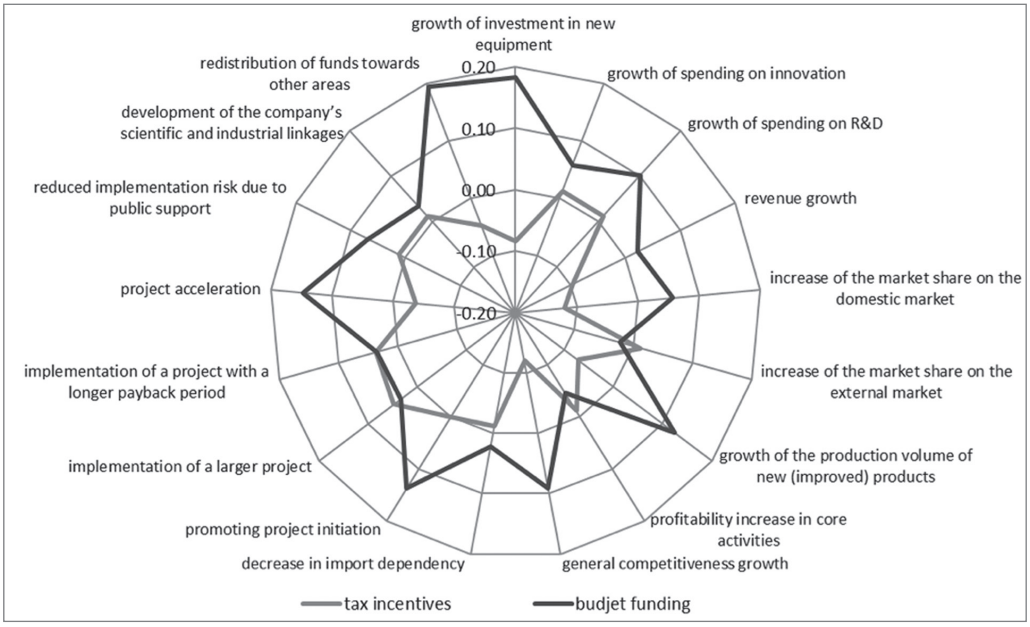
The main problem here is that the indicator is unobservable, so it is necessary for the calculation to find an approximation. Steps 1 and 2 enabled us to do this by using the observed values of this indicator obtained for the "most similar" companies, which were not recipients of a specific support type. As a result, the functional dependence takes the following form:

$$ATT_{ij} = E(Y_{ij}^1 \| S_i = 1) \square E(Y_{ij}^0 \| S_i = 0) \quad (2)$$

Note that since the variables and are binary, the mean of ATTs are located in the range from - 1 to 1. The ATT's zero value corresponds to the case where the average values of indicators $Y_{ij}^1 \| S_i = 1$ and $Y_{ij}^0 \| S_i = 0$ (or the value of effect for recipient and non-recipient consequently) is identical. This means that the use of the specific instrument does not increase or reduce the likelihood of a certain effect compared with the total set of public incentive mechanisms. Thus, the ATT indicator in this case reflects the "relative" additionality, provided by a certain dimension (tax or financial) on the background of general public support policy.

Evaluation results indicate that across almost all considered effects the impact of financial measures exceeds the impact of tax incentives. Exceptions are only two output indicators (the market share on the external market and profitability) and the scale and scope additionality (Figure 5).

Figure 5 – Relative additionality of tax and financial support – average treatment effects on the treated sub-sample of companies which received government support in 2013-15



Sources: prepared by the authors

2 A similar approach is used, for instance, in studies Fier et al. (2006); Baghana (2010); Marzucchi, Montresor (2013); Cantner, Kosters (2015).

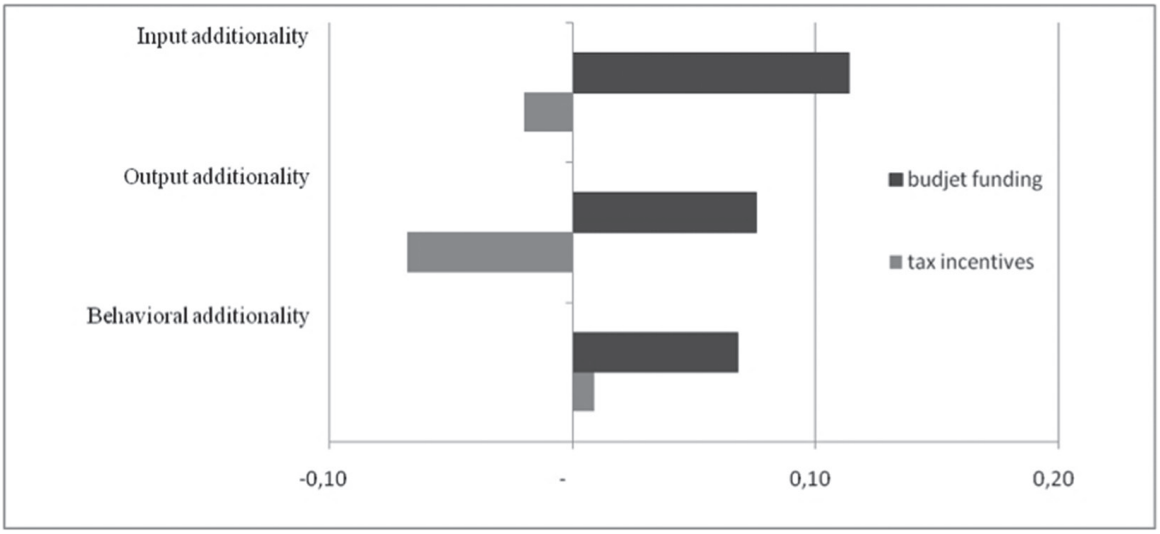
The most significant "failure" of the tax instruments' impact is observed in respect of the investments in new equipment, the market share on the domestic market and the general competitiveness of the company. At the same time, it is important to note that the crowding out effect is inherent to a much greater extent to financial instruments than to tax measures.

Aggregation of ATT values for input, output and behavioral additionality revealed that financial support most strongly affects the input company characteristics with a noticeable positive impact also on its output and behavioral parameters (Figure 6). In comparison tax measures have lower additionality effects, especially in terms of output additionality.

We take now a brief look at the question how company characteristics relate to specific effects. We calculated the parameters of ordinal regression models in which the dependent variables were the individual treatment effects on the treated companies[3] - the differences between the effects (with values 1 or 0) in the "recipient non-recipient" pairs: - . Explanatory variables reflect all of the above "standard" characteristics of companies (Table 3), except for the industry sector which was not sufficiently representative in the sub-sample formed by pairs of recipients and non-recipients of tax and financial support. For this reason, the industries have been aggregated according to their technology level[4]. In these models values of control variables correspond to those of the recipients (treated). Below only results for the meaningful models are presented (Table 5).

The results though should be interpreted with great caution due to the relatively small number of observations (pairs of "recipient / non-recipient"). Nevertheless, we believe it is important to note two consistent patterns (at least as a hypothesis for further empirical testing on larger samples): first, the effects of tax support occur more often in companies with state or mixed ownership and, second, the positive impact of tax incentives is less typical for small businesses than for larger firms.

Figure 6 – Relative input, output and behavioral additionality of tax and financial support measures – average effects on the treated (ATT) sub-sample of companies which received government support in 2013-15



Sources: prepared by the authors

3 This approach is used, in particular, in Hottenrott, Lopes-Bento (2013).

4 In our sample the low-tech industries include textiles, clothing and footwear, wood processing and pulp and paper industry, medium-tech includes metals and fabricated metal products, high-tech includes chemical industry and mechanical engineering (for more details see Zudin (2015)).

Table 5. Individual treatment effects on the treated for tax and funding instruments - calculation results of the binary logistic regression model

		Effects of tax incentives							Effects of public funding			
		increase in investment in new equipment	increase in volume of spending on innovation	growth of revenue	increase in general competitiveness	launch of a new project	implementation acceleration of the project	project implementation risks decline	increase in investment in new equipment	increase in volume of spending on innovation	network development	project implementation risks decline
Technology level of sector	low			+ *				+ *		- **		
	medium	control										
	high		+ **		- *	- **						
Operation period	less than 5 years				- ***				+ **			
	5-10 years				- *			- *		- **		
	10-20 years	control										
	more than 20 years	+ *			- *					- **		
Ownership	State and mixed		+ ***		+ **		+ **	- ***	+ ***	+ ***		+ *
	private	control										
Number of employees	less than 100 emp.				- ***	- ***			- *			
	101-200 emp..						+ *					
	201-500 emp..	control										
	more than 500 emp..	+ ***	+ *									
Financial condition	poor							+ **				
	satisfactory	control										
	good		+ **								- *	- *
Chi-square		18,05*	19,36*	17,60*	24,46**	19,92**	18,63*	20,74**	32,05***	28,43***	18, 84*	21,11**
N		83							88			
Maximum VIF value		2,55							5,13			

Notes:

* – significant a 10% level;

** – significant a 5% level;

*** – significant a 1% level.

Sources: prepared by the authors

DISCUSSION AND POLICY IMPLICATIONS

1. The ongoing Russian public policy to stimulate firms' development through a substantial variety of instruments for different target groups (e.g. Kuzyk, Simachev (2013))[5] is characterized by a strong emphasis on sufficiently large and long-operating companies. Such a result is not surprising, especially not for the Russian economy. Positive relationship between the size of the firms and the likelihood of receiving government support has been identified in a number of empirical studies (e.g. Fier, Heneric (2005); Aschhoff (2010); Simachev et al. (2014a)).

The focus shift of public support towards large and long operating companies occurs due to a number of factors. Firstly, these companies are more "visible" for the state and objectively are better able to lobby their interests in the government. Secondly, a large established business has a strong and highly diversified system of connections with the public authorities and a lot of experience in attracting and using government support. The latter is particularly important because, as has been shown in several studies, the company which has previously received support is more likely to receive it in the future too[6] (e.g. Falk (2006); Aschhoff (2009)). Thirdly, as it is noted in the study of Garcia and Monhen (2010), a greater proportion of large companies in the set of public support recipients may indicate a risk aversion of the government: indeed, support of a relatively small number of large companies in comparison with a lot of smaller firms is associated with lower transactional costs of support allocation and administration. Moreover, because large companies often demonstrate a "formally" higher innovation activity authorities tend to support them to generate a pseudo-positive result important for reporting (Simachev et al. 2014a). Finally, especially in periods of crisis, the government is more inclined to support large companies across particular industries, regions or/and the national economy as a whole[7](Simachev et al. 2010).

The question on the relative efficiency of government support for small and large firms is rather controversial. Today one can find empirical evidence of significant influence of government support both on SMEs, including behavioral changes (Loof, Heshmati 2005; Wanzenbock et al. 2013), as well as substantial corresponding changes in large firms (Falk 2006). The obtained results of our study rather confirm the second point of view. However, due to the relatively small number of relevant observations, we can only hypothesize that in Russia instruments of government support (especially tax incentives) provide positive changes mainly for medium and large sized firms.

Our view is that "quality" of the recipients and not formal characteristics of beneficiaries (such as size, age etc.) should matter and qualify for government support. The recipients of government support should have a big potential for further successful development and, what is more important, demonstrate abilities to implement it. However, in Russian realities that principle is not always followed. In periods of relative economic stability the government mostly supports successfully developing firms (e.g. Simachev et al. (2014a)), whereas crises force the government to shift the support focus towards troubled companies, especially if these are of a great importance in the context of providing socio-economic stability in the region or/and in the whole country (Higher School of Economics, Interdepartmental Analytical Center 2009; Mau 2010).

2. Tax and financial instruments of government support have de facto different target audiences: the use of tax incentives is not likely for small firms, whereas medium-sized companies relatively rarely appear to be recipients of financial support. The former can be the reflection of both the imperfect parameters of the tax instruments (their rate, base, etc.) for small businesses and the existence of significant implementation and administration problems, which are acceptable for large companies but too excessive for small firms. The fact of relatively rare financial support of medium-sized firms can be considered as another empirical evidence of a lack of instruments aimed at funding medium-sized projects and companies. (see also Simachev et al. (2012)).

3. The relatively small impact of government support on science-business cooperation seems to us quite unexpected (abroad, this effect is among the most frequently observed ones, especially in the case of financial support – e.g. Pegler (2005); Busom, Fernandez Ribas (2008); Idea Consult (2009); Marzucchi, Montresor (2013)). This fact is rather discouraging, as the Russian government makes considerable effort to enhance linkages and interactions between the R&D sector and industry. In the last few years, the government initiated a number of policies fully or to a large extent focused on the development of cooperation: financial support for projects to develop high-tech industries, executed by companies in cooperation with universities and research institutions; creation of a technological platform network; approval and implementation by the largest public sector companies of the medium-term innovative development programs which include cooperation activities with universities and research institutions; promotion of support programs for the development of innovative territorial clusters etc. The absence of an explicit result of these efforts, to our mind, can be explained by the fact that government support often does not lead to the creation of new linkages and partnerships but only contributes to the "capitalization" of long-established ones (Simachev et al. 2014c). It is worthwhile to note that a significant contribution of government support to the improvement of existing science-business linkages and partnerships has been widely observed abroad (e.g. Georghiou et al. (2005); Lohmann (2014)).

4. Our empirical analysis as well as a significant number of earlier studies confirmed the importance of the financial support in providing all major aspects of additionality. Based on our results, we can say that in Russia the effects of the financial instruments cover all three main types of additionality. The main input effect is the increase of investment in new equipment; the main output effect is the increase of production of new and improved products, and the main behavioral effect is the initiation of new perspective projects and an acceleration of project

5 Despite the fact that Russian industrial innovation policy toolbox is rather diversified

6 In literature this effect is commonly called the Matthew effect: this term is used in the broader context with respect to scientific recognition (Merton 1968) and in a narrow sense in relation to public support for innovation (e.g. Crespi, Antonelli (2011)).

7 It is appropriate to mention the renewal in 2015 of the practice of public guarantee support provided for strategic organizations - the largest entities that have a significant influence on the formation of GDP, employment and social stability. Previously the government has resorted to such measures in the most acute phase of the previous crisis - in 2009 (Simachev, Kuzyk 2010).

implementation. It should be noted that project additionality (government contribution to firms' launching new projects) is one of the most frequently observed behavioral changes (e.g. Falk (2007); OECD (2006); Idea Consult (2009)). This cannot be said about acceleration additionality (when government support speeds up the course of the project) which, according to other research, occurs on a considerably smaller scale (see Figure 2).

Unlike financial instruments, tax incentives almost do not provide significant results in terms of additionality. The most considerable "failure" is observed in relation to such expected – but not realized - effects as the supposed increase in the firms' competitiveness, the supposed growing domestic market share and the supposed increase of investment in new equipment. The negative results concerning the last indicator seem quite surprising to us as a large set of tax incentives in Russia is basically intended to stimulate firms' investment activities. At the same time and in contrast to a number of foreign studies, which examined a significant impact of tax incentives on input characteristics related to innovation activity, first of all R&D expenses (e.g. Lokshin, Mohnen (2012); Bodas Freitas et al. (2015)), we cannot see tangible input additionality of such measures in Russia. Slightly noticeable additionality effect of tax instruments relate to scale and scope additionality (the growth of investment in ongoing projects and the increase of the acceptable payback period). It is worthwhile to note that the positive impact of tax support on scale and scope of ongoing projects, in contrast to initiation of the new ones, was rather often identified in economic literature (Guellec, Van Pottelsberghe 2003; Jaumotte, Pain 2005; Simachev et al. 2014b).

The identified clear dominance of financial instruments over tax incentives with regard to most additionality effects should, in our opinion, not be considered as an exhaustive evidence of the inefficiency of tax measures and even more not as a robust argument in favor of abandonment of this element within the innovation policy portfolio. Indeed, the set of tax instruments obtains a number of important advantages. Actually, they are potentially available for a wider range of recipients than direct funding instruments. Moreover, they are associated with lower implementation and administration costs (Simachev et al. 2014b), they do not involve government intervention in market mechanisms and, what is important, they are not directly linked to budget allocation processes (Gokhberg et al. 2014). It is also important that tax measures and public funding instruments have substantially different beneficiaries. Finally, tax incentives produce to a noticeably less degree crowding out effects (replacement of private funds by public ones - e.g. David et al. (2000); Jaumotte, Pain (2005)) which is confirmed by the results of our study.

This situation, in our opinion, is explained by the fact that the expected benefits from the tax breaks are taken into account by firms ex-ante - when making a decision on the initiation of projects and defining their parameters. Financial support, on the contrary, is often only a possible, but not a guaranteed option, so in the case of obtaining such support companies prefer not to increase the project funding due to already fixed project parameters and redistribute funds for other needs.

In addition, budget funds are often invested in obviously successful projects that would have been carried out without public support on the grounds of the above-noted tendency of risk aversion of public authorities motivated by the need to demonstrate high efficiency in their programs (David et al. 2000; Klette et al. 2000; Wallsten 2000; Lach 2002).

The arguments given above, however, do not mean that tax incentives do not need improvement. On the contrary, the revealed significant "gap" in the effectiveness between financial and tax support shows in our opinion that the latter is just more in need of improvement - in terms of input and output effects, which are often observed abroad (Hægeland, Møen 2007; Lokshin, Mohnen 2012; Bodas Freitas et al 2015; Montmartin, Herrera 2015), but almost cannot be traced in Russia.

However, tax incentives productivity should not be improved through its "enrichment" with features and attributes of financial mechanisms as this would eliminate the key beneficial characteristics of tax incentives, which are the availability for a wide range of companies and low costs of use and administration. Thus, the relatively recent "improvement" of a certain tax benefit (the ability to write off a given amount of R&D expenditure) resulted in the requirement for the companies to submit to tax authorities the full research reports, which increased the application costs of this tax break and, as a consequence, led to a sharp reduction in its popularity among firms (Simachev, Kuzyk 2015).

Finally, we would like to note that our evaluation results on the influence of tax and financial policies on companies need to be interpreted with caution. Strictly speaking, no study of this kind can claim to be a universal truth. Indeed, as evidenced by the results of numerous foreign empirical studies, even very similar mechanisms of public support can lead to significantly different results in different countries and over different periods of time, and this is due not only to the differences in the "design" of support tools, but also because of their high impact heterogeneity over sectors, companies' parameters and market functioning characteristics. The observed effects vary considerably over time, and some appear only with considerable lags.

Against the background of an objectively limited value of any single empirical research it is particularly important to reflect the huge number of studies devoted to the evaluation of impact of public support on companies, a great portion of which for the last fifteen years has been based on the concept of additionality. Such studies, especially if regularly carried out on the basis of statistical data for long observation periods, often include cross-country comparisons of the results. Finally, and most important - they are in demand by the government and are implemented in the decision-making system. For Russia, surprisingly, our attempt to estimate the additionality of the tax and financial support was perhaps the first one.

At the very end, we would like to emphasize the urgent need in Russia of introducing the practice of regular and independent assessments of the effectiveness of public policies. At the same time, in our opinion, one should focus on additionality effects caused by public support, which would not have occurred in its absence. Along with an estimation of input and output effects, it is essential to take also behavioral changes that determine to a large extent the stability of the public support impact on the companies into account. This would help to create the required information basis for decision making on public incentive policies (both existing and to be initiated in the future), and would contribute to learning and scaling of the best practices.

REFERENCES

Antonioli, D., Marzucchi A., Montresor, S. (2012) 'Regional innovation policy and innovative behaviours. A propensity score matching evaluation', Ingenio Working Paper, № 5.

Aralicaa, Z., BotriĐ, V. (2013) 'Evaluation of research and development tax incentives sheme in Croatia', Ekonomska Istrazivanja, Economic Research, 26(3), pp. 63-80.

Aschhoff, B. (2009) 'The effect of subsidies on R&D investment and success: do subsidy history and size matter?' ZEW Discussion Paper, 09-032, Mannheim.

Aschhoff, B. (2010) 'Who gets the money? The dynamics of R&D project subsidies in Germany', Journal of Economics and Statistics, vol. 230(5), pp. 522–46.

Bach, L., Matt, M. (2002) 'Rationale for Science and Technology Policy'. In: Georghiou, L.,Rigby, J., (Eds.) Assessing the Socio-Economic Impacts of the Framework Programme, Report to DG Research.

Baghana, R. (2010) 'Public R&D Subsidies and Productivity: Evidence from firm level data in Quebec', United Nations University - Maastricht Economic and social Research and training centre on Innovation and Technology, Working paper, August 2010.

Benavente, J., Crespi, G., Maffiolly, A. (2007) 'Public Support to Firm-Level Innovation: The Evaluation of the FONTEC Program', OVE Working Paper, 0507, Inter-American Development Bank, Office of Evaluation and Oversight (OVE).

Bodas Freitas, I., Castellacci, F., Fontana, R., Malerba, F., Vezzulli, A. (2015) 'The additionality effects of R&D tax credits across sectors: A cross-country microeconometric analysis', Centre for Technology, Innovation and Culture, University of Oslo, Working paper, March 2015

Buisseret, T., Cameron, H., Georghiou, L. (1995) 'What Difference Does It Make - Additionality in The Public Support Of R&D In Large Firms', International Journal of Technology Management, № 10.

Busom, I, Fernandez Ribas, A. (2008) 'The impact of firm participation in R&D programmes on R&D partnerships', Research Policy, vol. 37(2), pp. 240–57.

Callejón, M., García-Quevedo, J. (2005) 'Public subsidies to business R&D: Do they stimulate private expenditures?', Environment and Planning C: Government and Policy, 23, pp. 279-93.

Cantner, U., Kösters, S. (2015) 'Public R&D support for newly founded firms – effects on patent activity and employment growth', Journal of Innovation Economics & Management, vol. 1(16), pp. 7-37.

Catozzella. A., Vivarelli. M. (2011) 'Beyond Additionality: Are Innovation Subsidies Counterproductive?', Discussion Paper, № 5746.

Chudnovsky, D., Lopez, A., Pupato, G. (2006) 'Innovation and productivity in developing countries: a study of Argentine manufacturing firms' behavior (1992-2001)', Research Policy, 35(2), pp. 266-88.

Clausen, T. (2009) 'Do subsidies have positive impacts on R&D and innovation activities at the firm level?', Structural Change and Economic Dynamics, 20, pp. 239-53.

Crespi, F., Antonelli, C. (2011) 'Matthew effects and R&D subsidies: knowledge cumulability in high-tech and low-tech industries', Departmental Working Papers of Economics, University Roma Tre.

Crespi, G., Maffioli A., Mohnen, P., Vazquez, G. (2011) 'Evaluating the Impact of Science, Technology and Innovation Programs: a Methodological Toolkit', SPD Working Papers, 1104, Inter-American Development Bank, Office of Strategic Planning and Development Effectiveness.

Czarnitzki, D., Licht, G. (2006) 'Additionality of public R&D grants in a transition economy: The case of Eastern Germany', Economics of Transition, 14, pp. 101-31.

David, P., Hall, B., Toole, A. (2000) 'Is Public R&D a Compliment or a Substitute for Private R&D? A Review of Econometric Evidence', Research Policy, vol. 29 (4), pp. 497-529.

Falk, R. (2004) 'Behavioural Additionality Effects of R&D Subsidies: Empirical Evidence from Austria', WIFO Working Paper. Available at: http://www.fep.up.pt/conferences/earie2005/cd_rom/Session%20IV/IV.B/Falk.pdf

Falk, R. (2006) 'Behavioural additionality in Austria's industrial research promotion fund (FFF)'. Chapter 3 in: Government R&D Funding and Company Behaviour: Measuring behavioural additionality, Paris, OECD, pp. 59-74.

Falk, R. (2007) 'Measuring the effects of public support schemes on Firms innovation activities', Research Policy, vol. 36(5), pp. 665–79.

Fier, A., Heneric, O. (2005) 'Public R&D Policy: The Right Turns of the Wrong Screw? The Case of the German Biotechnology Industry', ZEW Discussion Papers, № 05-60.

Fier, A., Aschhoff, B., Löhlein, H. (2006) 'Detecting Behavioural Additionality: An Empirical Study on the Impact of Public R&D Funding on Firms' Cooperative Behaviour in Germany', ZEW Discussion Papers, № 06-037.

Garcia, A., Mohnen, P. (2010) 'Impact of Government Support on R&D and innovation', UNU-MERIT Working Paper Series, № 2010-034, United Nations University - Maastricht Economic and Social Research Institute on Innovation and Technology.

Georghiou, L. (1997) 'Issues in the evaluation of innovation and technology policy'. Chapter 3 in: Policy evaluation in innovation and technology: towards best practice, OECD.

Georghiou, L. (2002) 'Impact and additionality of innovation policy', IWT-Studies, №. 40, pp. 57-67. Available at: http://www.iwt.be/sites/default/files/english/files/iwt_studie40.pdf

Georghiou, L., Clarysse, B. (2006) 'Introduction and Synthesis'. Chapter 1 in: Government R&D Funding and Company Behaviour: Measuring behavioural additionality, OECD Publishing, Paris.

Georghiou, L., Malik, K., Cameron, H. (2005). 'DTI Exploratory study on behavioural additionality', PREST, Manchester Business School and University of Manchester, Working paper.

Georghiou, L., Rigby, J., Cameron, H. (eds.) (2002) 'Assessing the Socio-Economic Impacts of the Framework Programme. Report to European Commission DG Research'. Available at: http://ec.europa.eu/research/evaluations/pdf/archive/other_reports_studies_and_documents/assessing_the_socio_economic_impacts_of_the_framework_programme_2002.pdf

Gelabert, L., Fosfuri, A., TribÓ, J.A. (2009) 'Does the effect of public support for r & d depend on the degree of appropriability?', Journal of Industrial Economics, № 57, pp. 736-67.

Gokhberg, L., Kitova, G., Roud, V. (2014) 'Tax Incentives for R&D and Innovation: Demand versus Effects', Foresight-Russia, vol. 8(3), pp. 18-41.

Gok, A. (2010) 'An Evolutionary Approach to Innovation Policy Evaluation: Behavioural Additionality and Organisational Routines', PhD Thesis, Manchester Institute of Innovation Research, The University of Manchester. Available at: http://www.innovation-policy.org.uk/compendium/reference/Default.aspx?referenceid=1026

Gok, A., Edler J. (2011) 'The Use of Behavioural Additionality in Innovation Policy-Making', MBS/MIoIR Working Paper, № 627, The University of Manchester.

González, X., Jaumandreu, J., Pazó, C. (2005) 'Barriers to innovation and subsidy effectiveness', RAND Journal of Economics, №36, pp. 930-50.

Guellec D., Van Pottlesberghe, B. (2003) 'The impact of public R&D expenditure on business R&D', Economics of Innovation and New Technologies, vol. 12(3), pp. 225-44.

Hægeland, T., Møen, J. (2007) 'Input additionality in the Norwegian R&D tax credit scheme', Statistics Norway Reports, № 2007/47. Available at: http://www.ssb.no/a/publikasjoner/pdf/rapp_200747/rapp_200747.pdf

Higher School of Economics, Interdepartmental Analytical Center. (2009), Evaluation of Anti-crisis Measures Supporting the Real Sector of the Russian Economy', Voprosy Ekonomiki, vol. 5, pp. 21-46.

Hsu, F.M., Horng, D.J., Hsueh, C.C. (2009) 'The effect of government-sponsored R&D programmes on additionality in recipient firms in Taiwan', Technovation, №29, pp. 204-17.

Hottenrott, H., Lopes-Bento, C. (2013) '(International) R&D collaboration and SMEs: The effectiveness of targeted public R&D support schemes', DICE Discussion Paper, № 121.

Hud, M., Hussinger, K. (2014) 'The Impact of R&D Subsidies During the Crisis', ZEW Discussion Paper, № 14-024.

Idea Consult (2006) 'Study and Evaluation of the Behavioral Additionality of R&D subsidies'. Available at: http://www.iwt.be/sites/default/files/english/files/final_report_BA_IWT-subsidies_10_11_2006.pdf

Idea Consult (2009) 'Does Europe change R&D-behaviour? Assessing the behavioural additionality of the Sixth Framework Programme'. Available at: https://ec.europa.eu/research/evaluations/pdf/archive/fp6-evidence-base/evaluation_studies_and_reports/evaluation_studies_and_reports_2009/assessing_the_behavioural_additionality_of_the_sixth_framework_programme.pdf

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

Klette, T., Moen, J., Griliches, Z. (2000) 'Do Subsidies to Commercial R&D Reduce Market Failures? Micro Econometric Evaluation Studies', Research Policy, 29, pp. 471-95.

Knockaert, M., Spithoven, A. (2009) 'When do firm-technology intermediary interactions result in cognitive capacity additionality?', Working Paper, D/2009/7012/61.

Kuzyk, M., Simachev, Yu. (2013) 'Russia's Innovation Promotion Policies: Their Evolution, Achievements, Problems and Lessons'. Chapter 6.4 in: Russian Economy in 2012. Trends and Outlooks (Issue 34) (eds. Sinelnikov-Murylev, S. G.), Moscow: Gaidar Institute Publishers, pp. 435-79.

Lach, S. (2002) 'Do R&D Subsidies Stimulate or Displace Private R&D? Evidence from Israel', Journal of Industrial Economics, L (4), pp. 369-90.

Lohmann, F. (2014) 'The Additionality Effects of Government Subsidies on R&D and Innovation Activities in the Aviation Industry. A Project Level Analysis', Master's thesis. Available at: http://essay.utwente.nl/64836/1/Lohmann_MA_MB.pdf

Lokshin, B., Mohnen, P. (2012) 'How effective are level-based R&D tax credits? Evidence from the Netherlands', Applied Economics, vol. 44(12), pp. 1527-38.

Loof, H., Heshmati, A. (2005) 'The Impact of Public Funding on Private R&D Investment: New Evidence from a Firm Level Innovation Study', MTT Discussion papers, № 3.2005.

Lucena, A., Afcha, S. (2014) 'Public Support for R&D, Knowledge Sourcing and Firm Innovation: Examining a Mediated Model with Evidence from the Manufacturing Industries', CENTRUM Católica's Working Paper Series, № 2014-06-0002.

Marzucchi, A., Montresor, S. (2013) 'The Multi-Dimensional Additionality of Innovation Policies: A Multi-Level Application to Italy and Spain', SPRU Working Paper Series, № 2013-04.

Mau, V. (2010) 'Economic Policy in 2009: Between the Crisis and Modernization', Voprosy Ekonomiki, vol. 2, pp. 4-25.

Merton, R.K. (1968) 'The Matthew effect in science', Science, № 159 (3810).

Montmartin, B., Herrera, M. (2015) 'Internal and external effects of R&D subsidies and fiscal incentives: Empirical evidence using spatial dynamic panel models', Research Policy, 44(5). pp. 1065–79.

Montmartin, B., Herrera, M., Massard. N. (2015) 'R&D policies in France: New evidence from a NUTS3 spatial analysis', Working Paper GAEL, № 2015-11.

Neicu, D., Teirlinck, P., Kelchtermans, S. (2015) 'Dipping in the policy mix: do R&D subsidies foster behavioral additionality effects of R&D tax credits?', Economics of Innovation and New Technology, pp. 1-22. Available at: https://lirias.kuleuven.be/bitstream/123456789/445812/1/MSI_1409.pdf

Newey, Whitney K. (2009) 'Treatment effects', Quantile, №. 6, pp. 15-23.

OECD (2006) 'Government R&D Funding and Company Behaviour, Measuring Behavioural Additionality'. Available at: http://browse.oecdbookshop.org/oecd/pdfs/product/9206041e.pdf

OECD (2015) 'OECD Science, Technology and Industry Scoreboard 2015: Innovation for growth and society', OECD Publishing, Paris.

Pegler, B. (2005) 'Behavioural Additionality in Australian Business R&D Grant Programs: A Pilot Study', Department of Industry, Tourism and Resources, Working paper.

Polt, W., Streicher, G. (2005) 'Trying to capture additionality in Framework Programme 5 - main findings', Science and Public Policy, 32(5), pp. 367-73.

Poot, T., Hertog, P., Grosfeld, T., Brouwer, E. (2003) 'Evaluation of a major Dutch Tax Credit Scheme (WBSO) aimed at promoting R&D', Discussion Paper. Available at: http://ftp.zew.de/pub/zew-docs/evaluationR%26D/EBrouwer.pdf

Rodrik, D. (2008) 'The New Development Economics: We Shall Experiment, but How Shall We Learn?', Harvard University, John F. Kennedy School of Government, Working Paper Series, № rwp08-055.

Roper, S., Hewitt-Dundas, N. (2012) 'Does additionality persist? A panel data investigation of the legacy effects of public support for innovation', Available at: http://druid8.sit.aau.dk/acc_papers/u7flocbh7rlj5s6uof3gkgquivxo.pdf

Rye, M. (2002) 'Evaluating the Impact of Public Support on Commercial Research and Development Projects: Are Verbal Reports of Additionality Reliable?', Evaluation, №8(2), pp. 227-48.

Simachev, Yu, Kuzyk, M. (2010) 'State guarantee support of strategic companies as Russian anti-crisis policy tool: specifics, problems and lessons for the future', Jekonomicheskaja politika, № 3, pp. 64-89.

Simachev, Yu, Kuzyk, M. (2015) 'State policy to promote scientific and industrial cooperation'. Chapter 6.4 in: Russian economy in 2014. Trends and outlooks. (Issue 36) (eds. Sinelnikov-Murylev, S. G.), Moscow: Gaidar Institute Publishers, 2015, pp.m465-511.

Simachev, Yu., Kuzyk, M., Feygina, V. (2014a) 'Fostering firms' innovation activities in Russia. Empirical analysis of beneficiaries and effects'. In Yasin, Ev.G (Ed) (2014) XIV April International Academic Conference on Economic and Social Development, vol. 3, pp. 369-88.

Simachev, Yu., Kuzyk, M., Feygina, V. (2014b) 'Public Support for Innovation in Russia: What Can We Say about Tax Incentives and Public Funding?', Russian Management Journal, vol. 12(1), pp. 7-38.

Simachev, Yu., Kuzyk, M., Feygina, V. (2014c) 'R&D Cooperation between Russian Firms and Research Organizations: Is There a Need for State Assistance?', Voprosy Ekonomiki, vol. 5, pp. 4-34.

Simachev, Yu., Kuzyk, M., Feygina V. (2015) 'Public Support for Innovation in Russian Firms: Looking for Improvements in Corporate Performance Quality', International Advances in Economic Research, № 21(1), pp. 13-31.

Simachev, Yu., Kuzyk, M., Ivanov, D. (2012) 'Instruments of government support for innovative companies in Russia: beneficiaries and issues'. In: Znaczenie innowacji dla konkurencyjności międzynarodowej gospodarki. (eds. Rynarzewski, T., Mińska-Struzik, E.), Poznan: Wydawnictwo Uniwersytetu Economicznego w Poznaniu, pp. 183-201.

Simachev, Yu., Kuzyk, M., Ivanov D. (2012) 'Russian Financial Development Institutions: Are We on the Right Way?', Voprosy Ekonomiki, vol. 7, pp. 4-29.

Simachev, Yu, Kuzyk, M., Kuznetsov, B. (2010) 'Assessing the impact of various anti-crisis measures on the firms in manufacturing industries', Jekonomicheskaja politika, №1, pp. 122-134.

Viljamaa, K., Piirainen, K., Kotiranta, A., Karhunen, H., Huovari, J. (2013) 'Impact of Tekes activities on productivity and renewal'. Available at: http://www.tekes.fi/globalassets/global/nyt/uutiset/productivity-and-renewal-2013.pdf

Wallsten, S. (2000) 'The Effects of Government-Industry R&D Programs on Private R&D: The Case of the Small Business Innovation Research Program', RAND Journal of Economics, 31, pp. 82-100.

Wanzenbock, I., Scherngell T., Fischer, M. (2013) 'How do firm characteristics affect behavioural additionalities of public R&D subsidies?', Technovation, vol. 33 (2-3), pp. 66-77.

Zudin, N. (2015) 'Relationship of the technology level of sector with company characteristics and public support', Innovations, №6, pp. 61-70.

AUTHORS

YURI SIMACHEV
National Research University Higher School of Economics
20 Myasnitskaya str., 101000 Moscow, Russian Federation

MIKHAIL KUZYK
Interdepartmental Analytical Center
P.O. Box 35, 31/29 Povarskaya str., 121069 Moscow, Russian Federation

corresponding author:
NIKOLAY ZUDIN
Center for Strategic Research
10 Vozdvizhenka str., 125009 Moscow, Russian Federation
T: +79165715843
E: N_zud@mail.ru

VALORIZING SSH RESEARCH: TOWARDS A NEW APPROACH TO EVALUATE SSH RESEARCH' VALUE FOR SOCIETY

IOANA GALLERON, MICHAEL OCHSNER, JACK SPAAPEN AND GEOFFREY WILLIAMS

In the last decades, we have witnessed a shift towards accountability and new public management practices in the management of universities in most countries. (Hamann, 2016; Hammarfelt and De Rijcke, 2015; Kekäle, 2002; Mali et al., 2016). Due to the pressure to be efficient and accountable, universities have implemented comprehensive evaluation procedures for research performance and research impact (Geuna & Martin, 2003). In addition, the availability of quantitative data and the preference of managers to use numbers to compare performance, has led to evaluation systems that are mostly based on science indicators, either drawing on data from Thomson Reuters' Web of Science or measuring direct, and sometimes indirect, economic effects. At first, this mainly affected the Science, Technology, Engineering and Mathematics (STEM) disciplines, but the budget constraints following the 2008 global financial crisis and its aftermath, as well as the continuing demand for accountability, led to wider implementation of quantitative research assessments, including the Social Sciences and Humanities (SSH) (Burrows, 2012; Guillory, 2005). However, while such campaigns led to considerable results in some cases, improving the overall performance of certain research systems, they were all confronted with numerous difficulties when it came to the evaluation of the SSH. A much broader resistance has developed against the quantitative approach of research quality and impact, also in the STEM-disciplines, as witnessed by the San Francisco Declaration on Research Assessment (DORA).

Research evaluation has always been perceived as a difficult area for the SSH for various reasons, amongst which being the wide variety of disciplines, approaches and practices brought together under the umbrella term of SSH. Problems mostly arise from the fact that the most common procedures have been fine-tuned to hard sciences and their production and communication practices, and as such they are ill-adapted to the research practices, to the national variations and to the dissemination traditions in the SSH disciplines. No wonder that a certain reluctance grew in the SSH fields, all the more so since SSH scholars believe strongly in the value of their disciplines for the advancement of knowledge, and in the contributions they can make to education, culture, the political system, work related issues, etc. As a consequence, the development of assessment procedures that are able to adequately review the work of SSH researchers have become necessity.

This paper endeavours to present the rationale for a valorising evaluation of SSH research. It will start with a presentation of the results of an initial survey about SSH research evaluation in Europe, conducted within the COST Action 15137 ENRESSH (the European Network for Research Evaluation in the Social Sciences and Humanities). The survey allowed us to capture perceptions about the main principles informing SSH research evaluation, to advance towards a typology of different evaluation models, and to better identify the problems linked to these models. In the second part, looking at good practices and based on these preliminary results, we will present an approach that combines performance and impact in a way that can represent a solution for SSH evaluation, and possibly beyond. In the third part, we will move to an analysis of the difficulties of valorising SSH through evaluation. These have been discussed to a certain extent in the recent literature (Spaapen & van Drooge, 2011; Molas-Gallart & Tang, 2011; de Jong et al., 2014), especially after different countries adopted evaluation campaigns introducing the criteria of "societal impact". However, while these discussions concerned questions such as attribution of impact to a specific research endeavour, or technical difficulties, such as how to document impact (Penfield et al., 2014), this paper will look more at the challenges of a fine understanding of the exact place of SSH in science and society, as a prerequisite for their evaluation.

I. TOWARDS A TYPOLOGY OF SSH RESEARCH EVALUATION SYSTEMS IN EUROPE

It is well known that evaluation practices differ widely across countries, and, over time, scholars have proposed different typologies of research evaluation systems (Coryn et al., 2007; Hicks, 2010; 2012; Martin & Geuna, 2001; 2003; von Tunzelmann & Mbula, 2003). However, none of the observations focused on SSH research evaluation in detail. Moreover, typologies have to date focused on a small number of countries, mostly those for which information on evaluation practices is available and widely discussed, such as the United Kingdom (RAE/REF), the "Norwegian system" (based on CRIStin), the evaluation in Belgium/Flanders (based on VABB_SHW), or the protocol for evaluation used in the Netherlands (SEP 2015-2021).

Given this lack of broad knowledge on SSH evaluation procedures, one of the first endeavours of ENRESSH was to observe and compare how research in the SSH is evaluated in different countries. The focus of the project is on European countries, even if, as the ENRESSH network expands, we start to be able to gather insights as to how the SSH are evaluated more widely.

To create a typology along which the countries can be classified according to their evaluation systems, a Delphi-like approach was adopted (for the use of the Delphi method to create a typology of evaluation systems, see Coryn et al., 2007; for a Delphi-method in the context of SSH research evaluation, see Hug, Ochsner & Daniel, 2014). The procedure consists of five steps. In a first step, a provisional typology was developed by the members of the Steering Committee and selected specialists from the Management Committee of the Action. In a second step, a survey based on this typology was administered to the specialists of the COST Action. The purpose was not at this stage to classify the countries, but rather to optimise the typology and to test the consistency of the classification among the respondents from the same country. The results and the feedback are being used in a third step to adapt the typology and to build an adapted questionnaire, that will be administered again to the specialists in a fourth step. Finally, the results as well as supplementary documents will be used to classify the evaluation systems of the countries. In the following, we are reporting results from the first two steps.

For the development of the initial typology, we started with a literature review, allowing us to identify several characterising dimensions on the basis of existing typologies. To these, we added some aspects we found were missing and/or specific to the SSH. As a result of this process, we designed a first typology consisting of the following dimensions: level of the evaluation protocol; differentiation; who is evaluating; funding; method; timeline; transparency; and costs. They are described in more detail as follows:

1. Evaluation is organised at different *levels* (von Tunzelmann & Mbula, 2003). Some countries have a national evaluation system, while in other countries, evaluations are organised at the regional level or subject to each university's autonomy. We differentiated between the level of organisation of the evaluation system on the one hand and the level on which data for evaluations are collected (existence of national, regional or institutional databases).
2. Research practices and communications in the SSH differ in a number of ways from research in the STEM disciplines. For example, commonly used evaluation practices, e.g. bibliometrics based on Web of Science data, are not readily applicable in the SSH (see, e.g., Hug, Ochsner & Daniel, 2013; Hicks, 2004; Nederhof, 2006; Ochsner, Hug & Daniel, 2012) and similar issues also arise with applied research (Furlong & Oancea, 2005). As a result, an additional dimension that is not yet present in the existing typologies[1] must be added: *differentiation*. It includes two aspects: a) whether there are specific methods or procedures to evaluate SSH research and b) whether there are different evaluation procedures for applied and for basic research.
3. Different *bodies* can be responsible for conducting or supervising the evaluation (Geuna & Martin, 2003; Hicks, 2010; 2012; von Tunzelmann & Mbula, 2003). Sometimes the differentiation between the level on which the evaluations are organised and the body responsible for evaluation is not very clear (von Tunzelmann & Mbula, 2003).
4. Evaluation can be linked to *funding* or serve formative reasons (Coryn, Hattie, Scriven & Hartmann, 2007; Geuna & Martin, 2001; 2003; von Tunzelmann & Mbula, 2003). Obviously, there is also the possibility that the evaluation outcome is not officially linked to funding but is nevertheless used for funding purposes by other institutions, or inside the evaluated institution.
5. Different *methods* can underlie the evaluation procedure (Coryn et al., 2007; Geuna & Martin, 2003; Hicks, 2010; 2012; von Tunzelmann & Mbula, 2003). This dimension has the following aspects: a) the principal method; b) whether and what kind of data is used; and c) criteria that are used if peers are involved.
6. Evaluations involve a time dimension. Two aspects are linked to time: a) evaluations can be repeated, thus the *time* of an evaluation cycle (in other words whether it is consistent and systematic) is a first aspect (Coryn et al., 2007; Hicks, 2010; 2012); and b) evaluations look back at a certain time window, which constitutes a second aspect (Hicks, 2010; 2012).
7. *Transparency* is an important dimension regarding dissemination and the use of the evaluation (Dahler-Larsen, 2012; Hammarfelt, Nelhans, Eklund & Astrom, 2016; Hicks, 2010). This is closely linked to the method applied and whether there is a link to funding. As results of evaluations can be seen as indicators of quality themselves, transparency and a reflective dissemination are crucial. Evaluations engage therefore an ethical responsibility (see Hicks et al., 2015; Klein, 2008); it is also a requirement for the construction of indicators in general (see the OECD Handbook of Composite Indicators, Nardo, Saisana, Saltelli & Tarantola, 2005), as well as in program evaluation (Morris, 2015). However, while Hicks mentions that "Most systems emphasise transparency of methods and data" (Hicks, 2010, p. 39), she does not use transparency for the typology. In our case, we use three aspects of transparency: a) the methods for calculating the final scores, when these are an outcome of the evaluation; b) the methods for linking scores to funding, if funding is linked to evaluation; and c) the publication of the results.
8. Evaluations come with a *cost* in both time and money. Hicks emphasises the need to include the cost of evaluation in a typology, but states that "cost is rarely discussed" (Hicks, 2010, p. 34). She observes also that assessing "costs and benefits […] is impossible" (Hicks, 2012). Geuna and Martin (2003) also raise the question of the cost/benefit ratio of performance-based research assessments. However, they did not investigate whether this was a topic in the countries. Rather, they argue that such systems, in general, will not have a positive cost/benefit ratio in the long run as the procedures become more and more complex and the returns on investment diminish as more countries apply the same procedures. We included two aspects regarding the costs in our typology: a) whether (estimated) costs are made public and b) whether there are efforts to estimate cost/benefit ratios.

1 Hicks (2010; 2012) mentions that field-specific approaches are necessary and states that "all systems are sensitive to differences in the patterns of fields' output" (Hicks, 2010, p. 37). In particular, the SSH are to be treated differently. However, she does not classify or specify how the systems account for differences.

A questionnaire based on these dimensions and aspects was then devised and administered, in March/April 2016, to the sixty members of the Management Committee (MC) of the COST-Action. The latter are all experienced in topics related to SSH research evaluation and represent their countries in the Action. The purpose of the survey was to get a first impression as to how the dimensions are used by the representatives of the countries to describe the evaluation system in their country.

Despite the time constraints – the fieldwork lasted for less than one month – , 43 persons from 25 countries filled in the questionnaire, of which 36 respondents from 22 countries answered all the questions. Countries were represented by one to five respondents; ten countries were represented by more than one respondent at least for a part of the questionnaire. The 25 countries in the study were: Austria, Belgium, Croatia, Czech Republic, Denmark, Estonia, France, Germany, Iceland, Ireland, Italy, Lithuania, Macedonia, Malta, Republic of Moldova, the Netherlands, Norway, Poland, Portugal, Romania, Serbia, Slovenia, Spain, Switzerland, and United Kingdom.

The results confirm that the dimensions of the existing typologies do not suffice to adequately describe SSH evaluation systems. First, there is variance between countries in the dimensions and aspects we added, e.g., differentiation, transparency and cost. Second, the intensive use of the comment fields showed that even more dimensions or aspects should be taken into account in order to adequately reflect the different evaluation systems. Obviously, this is also due to the more heterogeneous selection of countries included in this study as opposed to the selections on which other existing typologies are based.

While there was agreement between representatives of the same countries regarding the *methods* applied, *who* is responsible for the evaluation and whether results are used for *funding* decisions, there was much disagreement regarding the other dimensions. This disagreement might be due to a number of reasons. For example, the comments showed that while the survey had set out to tackle national evaluation systems regarding ex-post research evaluation, the respondents had all kinds of evaluations in mind, from ex-ante evaluations of research proposals to appointments to professorships and ex-post evaluations. Some also mentioned that they differentiate between evaluation and assessment, in terms of defining evaluation as formative and assessment as linked to funding. While this might be due to an inadequate definition of the terms in the survey, we rather interpret this as reflecting the national differences in the organisation of evaluation. For instance, in some countries appointments to professorships are organised nationally, and thus were included in the responses to the survey by the representatives of these countries, while in others, appointments are organised at the institutional level, and thus not subject to this survey for the representatives of those countries. Bearing this in mind, rather than being restrictive in our definitions, we are gradually adapting the dimensions and aspects of the typology, so as to take into account these national differences.

Besides these insights into the national differences of evaluation systems, the questionnaire confirms the existence of an accountability-based evaluation applied to the SSH in many countries. More often than not, evaluations are national: in 19 out of the 25 countries covered by the survey, respondents report a form of national evaluation, whether it be institutional or individual, and this proportion remains high even if we exclude the three cases with a strong disagreement between respondents from the same country (Belgium, Croatia and Spain). 20 respondents from 13 countries affirm evaluation is related to funding; these figures become 25 and 15 when one adds those respondents considering that "officially, evaluation is used to provide feedback (formative evaluation), but funders or universities base their funding decisions on evaluation outcome". This is to be corroborated with the answers from six countries according to which evaluation is "solely for feedback", as well as with the fact that the degree of agreement between respondents from the same country is higher than for other questions.

Interestingly enough, the situation is ambiguous regarding the differentiation dimension. Only 18 out of 43 respondents from 14 out of 24 countries affirm the exercises to be adapted to the SSH; four respondents perceive an evaluation to be SSH-specific if no citation data is used for some of the SSH, but affirm that, otherwise, the same procedures are applied. This is to be corroborated with the fact that only one scholar from one country affirmed that there is no use of citation data for evaluating SSH research, an answer one would have expected more often linked to the answer "evaluation of SSH disciplines is SSH specific". In short, one gets the impression that the evaluation of the SSH is not always SSH-specific, even though research on evaluation strongly encourages discipline-specific procedures. Furthermore, if the evaluation is SSH-specific, it is so because of the failures of existent procedures (e.g. bibliometrics cannot be applied to the SSH), rather than because it was carefully designed to reflect SSH research practices and goals.

2. FROM ACCOUNTABILITY TO VALORISATION

Is it possible to overcome the shortcomings of the existing evaluation protocols applied to the SSH, as observable through the responses quoted above, and to propose an intellectual frame, as well as methods and techniques truly adapted to these disciplines? The way forward seems to be a shift from the main principles, frameworks and practices of current research evaluation, as schematically described above, towards an approach looking for a combination of performance and valorisation of research in these disciplines. This does not mean pleading for a one-size-fits-all approach, nor abandoning the criteria of scientific quality, and even less forgetting about the accountability of sciences to the society. The idea is to reorient the evaluation exercises in a way that would be both more acceptable to the SSH scholars themselves and also able to provide a much needed evidence for informed decision making. This has the advantage of allowing specialists in the research evaluation to concentrate upon the numerous questions that arise, instead of trying to adapt "traditional" evaluation methods, often metrics based, to the specificities of SSH research, a somewhat procrustean endeavour.

A valorisation model for evaluation starts from the assumption that SSH research produces value, both for academia and for society, and that a large part of this value is not measurable in quantitative terms, nor assessable in other tangible terms. SSH research often regards new perspectives and insights that may influence the organisation and structure of processes and sectors in society. Whether regarding the "hard" sciences or, in more recent analysis, regarding SSH research, it has been repeatedly demonstrated that "impact" does not repose on a linear model, and that major innovations, be they technological, economic or societal, are multifactorial and cannot be related with certitude to a specific research project, publication or team (Greenlagh et al, 2016; Bornmann, 2012; Bornmann, 2013).

Nevertheless, the assumption that the SSH produces value is far from

being a matter of pure belief. Recent developments have shown, one more time, that there is a continuum between "hard" and "soft" sciences (Desmond Hellman, 2016), and that underfunding or undervaluing of the latter may hinder important developments in the first when they are much needed (Bod, 2013). Also, what education brings to society cannot be easily measured, but there is much evidence that education in all kind of subjects (even in "obscure" disciplines and fields of research, such as rare and ancient languages, for instance), and not mere training in immediately employable, job market needed tasks, is the basis of an articulated democracy (see Nussbaum, 2010).

Consequently, a valorisation model should concentrate less upon the "value for money" dimension and more upon finding the ways to stimulate the production and the dissemination of SSH knowledge. Considering that the uptake of research advancements is uncertain and not programmable, a model should pay less attention to "impact" understood as "modification in B due to A", and more to the collaborative dimension in these disciplines, even if this means inviting many of them, traditionally characterized by a solitary hermeneutical approach, to a considerable epistemological shift. If the goal is to get the most from SSH research, a valorisation model should set criteria and standards by stimulating strongly connected SSH, both to academia and to society. In other words, evaluating to valorise involves understanding and rewarding high quality, interdisciplinary and societally connected research, rather than concentrating on either academic or societal impactful research. This has the advantage of evaluating scholars, teams or institutions on the basis of what they actually do (or not), including the pro-active and innovative ways they develop to engage with the scientific community and society, rather than on the basis of what the scientific community and/ or the society does (or not) with their research. While societal and academic relevance should always be pointed out for any research undertaken, it is important to understand that the actual impact cannot be demonstrated in an unrealistically short time-frame and using questionable evidence. Most impacts that really make a difference may take 10 to 15 years. This understanding may also help to prevent perverse effects such as focusing only on research that comes with low risk and with short-term attention in academia and society. At the same time, it may also result in a certain slowing of the race to publication and citation, and may alleviate the burden of collecting "proofs of impact" which weighs heavily on researchers from certain countries, to the detriment of the time they can actually dedicate to research and teaching itself. Having said that, it does make sense to look at short and medium term effects in the context of the larger innovation process, for example via contributions of researchers to that innovation process.

Fortunately, some large-scale experimentations of a valorising evaluation model have been conducted and are leading to an assessment of (SSH) research in more understanding ways, with both the scientific quality and the societal relevance assessed. In the literature, many examples of new approaches for evaluation of societal impact can be seen (see special issue of *Research Evaluation*, September 2011; RAND 2013; several HEFCE reports; Lyall 2013). Practical examples can be found in the Netherlands, where a comprehensive framework specifically for humanities research has been presented (https://www.qrih.nl/nl/)[2]. More methods around impact pathways are currently developed in the second working group of ENRESSH.

In short, novel approaches to evaluation need to include the following:

- Knowledge about SSH research production;
- Some form of socially distributed responsibility, stakeholder involvement;
- Focus on the context of application of knowledge, next to scientific excellence;
- Be subject to multiple accountabilities (collegial/professional vs. managerial).

3. DIFFICULTIES OF VALORISING EVALUATION OF SSH RESEARCH

INTRODUCING A VALORISING MODEL DOES NOT GO WITHOUT ITS OWN DIFFICULTIES.

While standards of quality are controversial in all disciplines, recent research shows that perceptions and conceptualisation of excellence are even more complex and fuzzy in the SSH (see, e.g., Furlong & Oancea, 2005; Hemlin, 1996; Ochsner, Hug & Daniel, 2014; Williams & Galleron, 2016). Also, while peer-review is generally universally acclaimed and accepted within this area, in many journals or publishing houses, as well as at other levels and institutions where evaluation is practiced by peers, procedures are far from being transparent and robust, and often have not been closely monitored or assessed against principles such as thoroughness and fairness (Hemlin, 2009). Moreover, the relationship between science and society is changing and evaluation mechanisms are bound to reflect that to a certain extent. Boundaries between the two spheres become blurred and stakeholders become more involved in collaborations with researchers. This means that in some cases their interests and goals have to be included in review systems, also in the peer review, or expert review as it becomes then (Hemlin, 2006).

Another difficulty is that "societal impact" as a concept is difficult to define since it depends heavily on the context. If we limit ourselves to SSH research, it is clear that a researcher doing work, for example, in the area of religious studies working on the integration of Muslims in Western societies will be working in a rather different context than a researcher who is working in history of technology aiming at a new curriculum for high schools. The former most likely has to collaborate with people from religious and other communities and with policy and law makers, while the latter might work in the context of secondary education. Debates will differ, and so will the needs of these different stakeholders. This affects the kind of products needed by stakeholders. In the case of religious studies, knowledge exchange and policy proposals might be a prime goal, in the case of history of technology, a course or a book might be the product. Regarding evaluation, this means that it is difficult to come up with societal impact measurements that are adequate for all or most fields in SSH. Moreover, while the word 'impact' has a linear connotation (with a sender and a receiver) these two examples confirm what has been suggested in the second section, that results can only be achieved in interaction with stakeholders, which has to be stimulated more than the actual impact being measured. A new course in history of technology cannot be developed without the school community that

2 The website is in Dutch, an English version will become available in September 2017

is supposed to work with it. And the knowledge developed by scholars who study religion can only help the debate about the migration crisis through a debate with other parties involved.

In discussing societal impact, much use is currently made of the quadruple helix model, in which government, industry, academia and civil participants work together "seamlessly". Such a biological metaphor raises several questions when put in the context of research evaluation. The biological double helix discovered by Watson and Crick only works because there is interaction between the strands, which are not just running in parallel (and keep working because there is RNA that detects and repairs flaws); the same applies to the quadruple helix model where it is vital to clarify what these strands are and what they seek to achieve. The first problem that arises is in deciding who stakeholders actually are, and then trying to identify their motivations, perceptions and goals as each of the current strands actually covers very different entities. The obvious stakeholders are the scholars themselves, but their objections against STEM geared evaluation approaches have rarely been taken seriously up until now. A second group of stakeholders is that of the policy makers and funders, which might be as heterogeneous as the SSH themselves. A third group is society, even more diverse: public organisations, NGO's, small and big industry, and the public at large, who mostly values the SSH for the cultural knowledge and wisdom that underlie stable democracies where freedom of thought is cherished. An analysis of all these strands is necessary to find out where and how they are connecting, and where they are not. Understanding the strands means raising awareness across the board so that common ground can be found. ENRESSH is already working towards this, with the aim to create a dialogue between different policy makers as well as opening up debate as to other aspects of the helix.

CONCLUSIONS

Evaluating to valorise is particularly important for the SSH disciplines. However, much research is still needed in order to proceed towards such a model. An in-depth understanding of SSH knowledge production processes and strategies is needed as a basis for developing evaluation procedures that adequately reflect the research practices, goals and aims of the SSH scholars. In parallel, the engagement of SSH researchers with societal challenges has to be attentively studied, so as to have a more comprehensive view of the ways in which interaction takes place in non-academic partnerships and environments of SSH research. Lastly, robust data about SSH production has to be gathered, and this means in many cases creating from scratch research information systems dedicated to SSH research outcomes.

ENRESSH seeks to accelerate progress on all these topics, through coordinating research projects going on in several European countries. While primarily aimed at reorienting the evaluation of SSH research, its results may prove useful for the entire of academia, as voices are numerous in the STEM sciences pointing out that this area is also diverse, that many disciplines are ill-served by a "one-size-fits-all" approach, and that evaluation driven by a narrow set of scientific excellence criteria and/ or demands of "usefulness" does not do justice to the wealth of contributions research is bringing to the advancement of knowledge and to the society.

REFERENCES

Bod, R. (2013). A New History of the Humanities. The Search for Principles and Patterns from Antiquity to the Present, Oxford, Oxford University Press.

Bornmann, L., (2012). "Measuring the societal impact of research", EMBO reports, vol. 13, no. 8.

Bornmann, L. (2013). "What Is Societal Impact of Research and How Can It Be Assessed? A Literature Survey", Journal of the American Society for information science and technology, 64(2), 217–233.

Burrows, R. (2012). "Living with the hÐindex? Metric assemblages in the contemporary academy", The Sociological Review, 60(2), 355–372, http://doi.org/10.1111/j.1467-954X.2012.02077.

Coryn, C. L. S., Hattie, J. A., Scriven, M., & Hartmann, D. J. (2007). "Models and Mechanisms for Evaluating Government-Funded Research: An International Comparison", in American Journal of Evaluation, 28(4), 437–457, http://doi.org/10.1177/1098214007308290.

Dahler-Larsen, P., (2012). "Evaluation as a Situational or a Universal good? Why Evaluability Assessment for Evaluation systems is a Good Idea, What it Might Look Like in Practice, and Why it is not Fashionable", Scandinavian Journal of Public Administration, 16/3: 29–46.

De Jong, S., Barker, K., Cox, D., Sveinsdottir, T., Van den Besselaar, P., (2014). "Understanding societal impact through productive interactions: ICT research as a case", in Research Evaluation, 23(2), 89–102.

De Jong, S., van Arensbergen, P., Daemen, F., van der Meulen, B., van den Besselaar, P., (2011). "Evaluation of research in context: an approach and two cases", in Research Evaluation, 20(1), 61-72.

Desmond-Hellman, Sue (2016), "The Latest Weapon in The Fight Against Disease Might Surprise You", published on October 26, 2016 on Linkedin.

Furlong, J., Oancea, A., (2005). "Assessing quality in applied and practice-based educational research: A framework for discussion". Oxford University Department of Educational Studies. Retrieved from http://www.esrc.ac.uk/ESRCInfoCentre/Images/assessing_quality_shortreport_tcm6-8232.pdf

Geuna, A., & Martin, B. R., (2001). "University Research Evaluation and Funding: An International Comparison", in SPRU Electronic Working Paper Series (Vol. 71).

Geuna, A., & Martin, B. R., (2003). "University Research Evaluation and Funding: An International Comparison", in Minerva, 41(4), 277–304. http://doi.org/10.1023/B:MINE.0000005155.70870.bd.

Guillory, J. (2005). "Valuing the humanities, evaluating scholarship" Profession (MLA), 11, 28–38.

Hamann, J. (2016). "The visible hand of research performance assessment", Higher Education, 1–19. http://doi.org/10.1007/s10734-015-9974-7.

Hammarfelt, B., & de Rijcke, S., (2014). "Accountability in context: effects of research evaluation systems on publication practices, disciplinary norms, and individual working routines in the faculty of Arts at Uppsala University", in Research Evaluation, 24(1), 63–77. http://doi.org/10.1093/reseval/rvu029.

Hammarfelt, B., Nelhans, G., Eklund, P., & Åström, F. (2016). "The heterogeneous landscape of bibliometric indicators: Evaluating models for allocating resources at Swedish universities", in Research Evaluation, http://doi.org/10.1093/reseval/rvv040.

Hemlin, S. (1993). "Scientific quality in the eyes of the scientist. A questionnaire study." Scientometrics, 27(1), 3–18.

Hemlin, Sven, Barlebo Rasmussen, Søren, (2006) "The Shift in Academic Quality Control", in Science Technology Human Values, vol. 31, no. 2, 173-198.

Hemlin, Sven, (2009) "Peer Review Agreement or Peer Review Disagreement: Which Is Better?", Journal of Psychology of Science and Technology, Volume 2, Number 1.

Hicks, D. (2004). "The four literatures of social science", in H. F. Moed, W. Glanzel, & U. Schmoch (Eds.), Handbook of quantitative science and technology research, Dordrecht: Kluwer Academic, 473–496.

Hicks, D. (2010). "Overview of models of performance-based research funding systems", in Performance-based Funding for Public Research in Tertiary Education Institutions, OECD Publishing, 23–52. http://doi.org/10.1787/9789264094611-4-en.

Hicks, D. (2012). "Performance-based university research funding systems", Research Policy, 41(2), 251–261. http://doi.org/10.1016/j.respol.2011.09.007.

Hicks, D., Wouters, P., Waltman, L., de Rijcke, S., Rafols, I., (2015). "The Leiden Manifesto for Research Metrics", Nature, vol. 520, 23 April 2015, p. 429-431

Hug, S. E., Ochsner, M., & Daniel, H.-D. (2014). "A framework to explore and develop criteria for assessing research quality in the humanities", in International Journal of Education Law and Policy, 10(1), 55–68.

Hug, S. E., Ochsner, M., & Daniel, H.-D. (2013). "Criteria for assessing research quality in the humanities: a Delphi study among scholars of English literature, German literature and art history", in Research Evaluation, 22(5), 369–383. http://doi.org/10.1093/reseval/rvt008

Kekäle, J. (2002). "Conceptions of quality in four different disciplines", in Tertiary Education and Management, 8(1), 65–80, http://doi.org/10.1023/A:1017947005085.

Klein, J. T. (2008). "Theoretical perspectives on team science, Evaluation of Interdisciplinary and Transdisciplinary Research: A Literature Review", American Journal of Preventive Medicine, Volume 35, Issue 2, Supplement, August 2008, p. S116-S123.

Mali, F., Pustovrh, T., Platinovsek, R, Kronegger, L, & Ferligoj, A. (2016). "The effects of funding and co-authorship on research performance in a small scientific community", Science and Public Policy, online first. doi:10.1093/scipol/scw076.

Molas-Gallart, J., Tang, P., (2011). "Tracing "productive interactions" to identify social impacts: An example from the Social Sciences", in Research Evaluation, 20(3), 219–226.

Morris, M. (2008). Evaluation ethics for best practices: cases and commentaries, New Yonk, The Guilford Press.

Nardo, M., Saisana, M., Saltelli, A., Tarantola, S., Hoffman, A., Giovannini, E., (2005). Handbook on constructing composite indicators. Methodology and user guide. Paris: OECD.

Nederhof, A. J. (2006). "Bibliometric monitoring of research performance in the social sciences and the humanities: A review", Scientometrics, 66(1), 81–100, http://doi.org/10.1007/s11192-006-0007-2.

Nussbaum, M. C. (2010). Not for Profit: Why Democracy Needs the Humanities. Princeton University Press: Princeton.

Ochsner, Michael, Hug, Sven, Daniel, Hans-Dieter, (2012). "Indicators for Research Quality in the Humanities: Opportunities and Limitations", Bibliometrie - Praxis Und Forschung, 1.

Ochsner, M., Hug, S. E., & Daniel, H.-D. (2014). "Setting the stage for the assessment of research quality in the humanities. Consolidating the results of four empirical studies." Zeitschrift für Erziehungswissenschaft, 17(6 Suppl.), 111–132. http://doi.org/10.1007/s11618-014-0576-4

Penfield, T., Baker, M.J., Scoble, R., Wykes, M.C. (2014). "Assessment, evaluations, and definitions of research impact: A review", in Research Evaluation, 23 (1), 21-32.

R**oyal Netherlands Academy of Arts and Sciences**. (2011). Quality indicators for research in the humanities. Amsterdam: Royal Netherlands Academy of Arts and Sciences. https://www.knaw. nl/shared/resources/actueel/publicaties/pdf/quality-indicators-for-research-in-the-humanities.

Spaapen, J., & van Drooge, L. (2011). "Introducing "productive interactions" in social impact assessment", in Research Evaluation, 20(3), 211–218.

Trisha Greenhalgh, James Raftery, Steve Hanney and Matthew Glover, (2016) "Research impact: a narrative review", in BMC Medicine, 14:78.

Tunzelmann, Von, N., & Mbula, E. K. (2003). Changes in research assessment practices in other countries since 1999: final report.

Williams, G., & Galleron, I. (2016). "Bottom Up from the Bottom: A New Outlook on Research Evaluation for the SSH in France." In M. Ochsner, S. E. Hug, & H.-D. Daniel (Eds.), Research Assessment in the Humanities. Towards Criteria and Procedures (pp. 181–198). Cham: Springer International Publishing. http://doi.org/10.1007/978-3-319-29016-4_14

AUTHORS

IOANA GALLERON
Paris III University, (France)
E: galleron@evalhum.eu

MICHAEL OCHSNER
ETH Zürich/ FORS, Lausanne (Switzerland)
E: ochsner@evalhum.eu

JACK SPAAPEN
KNAW (the Netherlands)
E: jack.spaapen@knaw.nl

GEOFFREY WILLIAMS
*University of Grenoble-Alpe*s (France)
E: williams@evalhum.eu

Acknowledgement: "This article is based upon work from COST Action ENRESSH (CA15137), supported by COST (European Cooperation in Science and Technology) http://www.cost.eu/ ".

SERVING VARIEGATED AUDIENCES: FROM RANKING ORIENTED EVALUATION TO MISSION ORIENTED EVALUATION

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

Academic researchers are under an ever growing pressure to demonstrate that the work they do not only has excellent scientific value but is also relevant to society's questions and challenges. Many governments have introduced targeted funding programs that demand societal impact of academic research, and so has the European Commission, in particular with the Grand Societal Challenges of the Horizon 2020 framework program. As a rule, such funding schemes expect academic researchers to team up with partners in society, depending on the topic of research coming from industry, the public sector or society at large. And this tendency to involve society in the setting of academic research agenda's may go even further witness the National Science Agenda introduced in 2014 by the Dutch government. In this program, all Dutch citizens were asked to submit questions they deemed worth to research.[1]

The consequence of this policy to steer academic research more into the direction of society and it's problems is that academic researchers have come to operate in a much wider context than the university context they were used to, and that they have to review the knowledge that they produce and the ways they communicate this with their newly emerging environment. This is not to say that before the changes in governmental funding policy academics did not interact with society, but both the scale and the mandatory character of the policy is very different than in the past. The societal challenges part of H2020 contains about 30 billion euros for the six year period between 2014 and 2020 (of the total budget of around 80 billion euro). Clearly, these changes also have consequences for the way research is evaluated; after all, the value of research needs to be assessed against a much wider context than the performance in the international literature. Next to excellence, societal impact then becomes an important criterion in research evaluation.

As Bornmann in his 2013 literature review shows, there is not a lot of consensus yet about what societal impact is or how to evaluate it. But much of the literature about research impact assessment stresses the importance of the network of societal stakeholders related to academic research (Bornmann 2013). Recent experiences in evaluation practices such as the British REF UK 2014 show considerable diversity of such networks, and of the ways to achieve impact (Manville et al., 2015). In this paper we will review the diversity of such networks for a number of research organizations inside academia, and the possible consequences for evaluation practices. We will argue that this diversity reflects the organizational and social characteristics of the specific network in which research organizations work, and as such requires a greater focus on these specific characteristics and on the mission of the units to be evaluated. This shift in evaluation focus implies a diminishing role for comparison with other units (unless they have the same mission), something that is an important goal of evaluation for many governors, managers and administrators who often are ranking oriented. But we think that the kind of evaluation we have in mind is more adequate for the societal networks in which research takes place these days. And we also think that evaluation still can be conducted in a systematic and robust manner. A shift from comparison oriented evaluations to the specific characteristics and mission of research units offers the opportunity to learn from and improve stakeholder relations, and thus, arguably, the impact of research. Mission oriented evaluations arguably help both the management of an institute and the (interaction with) the environment.

THE SOCIETAL NETWORK OF INNOVATIVE RESEARCH

Research that helps innovate society, whether this is through new technologies or new forms of organisation, new insights or processes, or otherwise, is mostly part of multi-actor endeavors, with participants coming from both science and society. This is certainly the case for the many research institutes that operate outside academia, institutes that have a specific mission for example in health research, environmental or energy studies. Typically, the research agendas of such institutes are developed in collaboration with relevant stakeholders, and the results of the research work is often published in media that have a wider reach than the academic community (Spaapen and Van Drooge 2011: 214).

But more and more, institutes operating within the academic sphere (universities, academy institutes, and other) also perform their research

1 The request of the Dutch government to come up with questions for the National Research Agenda resulted in close to 12,000 questions by the general Dutch public and a vast number of public and private organisations. After an intricate procedure led by the Dutch Academy of Arts and Sciences the agenda now presents 140 overarching scientific questions. The plan is that Dutch scientific research focuses on these in the coming years. (http://www.wetenschapsagenda.nl/?lang=en)

in the context of issues and challenges in policy or society at large. The networks that are formed by these wider communities are as a rule more diverse and often also more temporary than traditional academic networks. Both researchers and policy makers have to review their perspective on this position, both in terms of how they interact with the broader environment (for example with regard to their research agenda) and in terms of how they assess the success of new forms of collaboration. The networks are characterized by a variety of scientific stakeholders (various disciplines) and stakeholders from society, be it industry, government or society at large. Somehow, all these different backgrounds, interests and work practices have to be attuned in new arrangements in which goals, performance and monitoring and evaluation have to be elaborated. Elsewhere, we have collaborated with many colleagues to analyse such networks and researched what the consequences could be for the evaluation (www.siampi.eu).[2] The analysis there focused on the different types of interactions that take place between the stakeholders in such networks of research and innovation: (1) Direct, in the sense of "personal" interactions that evolve around face-to-face encounters, or through phone, email or videoconferencing; (2) Indirect interactions through some kind of material "carrier": these include texts such as policy reports, protocols, books, music scores and questionnaires as well as artefacts such as websites, software, exhibitions, devices; and (3) Material interactions occur when potential stakeholders engage in a financial contribution, a contribution "in kind," or when facilities are shared (Spaapen and Van Drooge 2011). To research these interactions SIAMPI used a variety of methods, most of them stakeholder oriented. Among them were face-to-face interviews with academics and societal stakeholders and focus groups.

In a number of experimental studies in the Netherlands ideas developed in the SIAMPI project were tested, with a particular focus on hybrid forms of output and on the composition of the stakeholder context.[3] Hybrid output is a product or performance based on robust scientific work directed towards a broader audience than fellow researchers. Most often, this regards documents like books or articles that are not published in the regular scientific journals (for articles) or as scientific monographs (for books). There is also a growing variation of hybrid output such as films, (serious) games, protocols. In this paper, we limit ourselves to the various forms of written output.

We researched this by using the method of contextual response analysis (CRA) to trace the uptake of hybrid output in the environment of academic research groups. The idea is to see whether this approach, that we will explain in more detail further on, can help institutes to present their wider relevance in a convincing way in the context of the Dutch national evaluation system, the Standard Evaluation Protocol (SEP 2015-2021). In preparation of the national research evaluations (every six years in the Netherlands) all academic research institutes in the Netherlands have to produce so-called self-evaluation reports in which they also address the wider context of research[4]. Also, a number of non-academic publicly funded research institutes, have taken the SEP as their evaluation model, often in adapted form.

The changing policy context is difficult for all academic researchers, but we noticed something interesting when talking to researchers in the areas of the social sciences and the humanities. Many of the fields in these areas, having difficulties with evaluation systems that primarily focus on bibliometric indicators, and on international comparison, saw in the shifting focus in evaluation discussions (to societal impact) an opportunity to develop assessment procedures that were more adequate to the work they did for variegated audiences. A lot of work in social sciences and humanities are oriented towards issues outside academia, serving variegated societal areas like health and wellbeing, politics, education, culture, migration, etc..[5] That kind of work is often not very visible in the international databases that underpin the majority of the bibliometric indicators. That is the main reason the gathered deans of the Dutch humanities faculties decided in 2015 to start a project in which they wanted to develop a humanities specific evaluation protocol that at the same time would fit into the ruling national SEP protocol. Part of this paper is based on work we are conducting for that project which is expected to finish by the end of 2016. Next to that, we conducted a study for a broad social science institute of the University of Amsterdam. This way, we were able to take a closer look at research in the humanities and social sciences. In this paper, we focus on our review of the hybrid output of both the social science institute which has a broad palette of research interests and of a number of Dutch research schools in the humanities. Research schools organize researchers from various universities on the basis of their research interest. Most of these schools contain a variation of (sub)disciplines.

VARIATION IN PUBLICATION PATTERNS AND STAKEHOLDER CONTEXT

What the social science research institute and the humanities research schools have in common is that they are academic, but at the same time strive to conduct research that is not only scientifically excellent but also to a certain extent addresses socially relevant issues. In other words, they adhere to the growing demand from the government to become more relevant for societal challenges and questions.[6] Research topics thus more and more have to be relevant for the scientific community and for society at large. And, all these institutes have to find

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

3 Studies included CPB Netherlands Bureau for Economic Policy Analysis, The Netherlands Institute for Social Research SCP, PBL Netherlands Environmental Assessment Agency, WODC Research and Documentation Centre of the Ministry of Security and Justice, Ministry of Foreign Affairs. More information can be found at www.adprins.nl and at the site of PBL for the following study: A.A.M. Prins, Contextual Response Analysis of reports of the Netherlands Environmental Assessment Agency PBL, Groningen 2012;

4 All publicly funded research in the Netherlands is evaluated once every six years via the so-called Standard Evaluation Protocol (SEP): http://www.vsnu.nl/sep

5 See for example Federation 2017

6 In its vision on science and science policy for the next decade, the Dutch government distinguishes three main themes: excellent research, maximum social impact and cradle for young talent (Wetenschapsvisie 2025, The Hague 2014

ways to involve stakeholders that are important for their work. These stakeholders vary from context to context, , and can vary from local, regional or national governments to schools, to hospitals to industry to NGO's or society at large. If such stakeholders are to be involved in developing the research agenda, it is necessary to know who are relevant stakeholders and what they are picking up from the research that is published. In the following, we will use a method, the Contextual Response Analysis (CRA) to gather information about the uptake of research output by the wider societal environment of an institute.

Arguably, the balance between the two goals of producing results for science and for society depends to a certain extent on the policy context in which these institutes operate, including reward systems and local incentives. Researchers in academia have been rewarded until now mainly for their contributions to the scientific debate, i.e. articles in international journals (for AISSR) and books (for the humanities faculties). Our approach aims at gaining more insight about the context by tracing output and getting information about the various stakeholders that are interested in the research produced by these institutes. The method has profited much from the general trend towards open access, which has made it much easier to trace variegated forms of output. As all of these institutes serve public goals, their output follows governmental (and in fact European) policy towards open access, and now as a rule is made publicly available, in print and via websites. The publications of the institutes we studied may thus reach different stakeholders both inside and outside academic circles and governments, which is pertinent to the evaluation of the innovation processes in which these institutes take part.

In the following we shortly present the social science institute and the humanities research schools. The Amsterdam Institute for Social Science Research (AISSR) unites all social science research of the University of Amsterdam. The research program focuses on the functioning of contemporary societies and their interrelationships from historical, comparative and empirical perspectives. The research program is organized into thematically focused groups with an anchor in one or more of the represented disciplines: sociology, geography, planning & development studies, political science and anthropology. According to its mission, AISSR aims to contribute to public debates on key issues – specifically contributing to interventions that address pressing societal problems – and to engage with relevant stakeholders.

In the case of AISSR the research we did was experimental in the sense that the management of this social science institute was interested to see whether our approach would give them new insights into the visibility of their output for a variety of stakeholders inside and outside academia. They agreed to participate in experimenting with our approach and use the outcomes in their respective self-evaluations. The study served to help both the development of our approach and to help the management of the institute to present themselves in a broader perspective related to their environment.[7] The five sections of AISSR came forward with a list of 127 hybrid publications.

The examples of the research schools in the humanities are taken from a broader exploratory study aiming at the development of indicators of quality and relevance of research in the Humanities in The Netherlands. The expected result of the exploratory study is a guide that helps researchers to work with the national protocol SEP 2015-2021 that in its current form was judged not fit for humanities research. The study, commissioned by the deans of all humanities faculties in The Netherlands, has invited the seventeen research schools to participate in the development of the indicators. The schools include a wide range of domains in the Humanities, including vested fields such as Archeology, Cultural History, Political History, Arts, Literature, and Theology and Religion Studies, and also multidisciplinary fields such as Cultural Studies or Media Studies.

The management of the humanities faculties we studied had a broader interest than the management of AISSR. They were looking for an approach that would help them to produce narratives for the national evaluation protocol SEP 2015-2021. So far, they found the protocol not fit for evaluation of the kind of research they conduct, and too much geared towards evaluation mechanisms used in the STEM fields. They were particularly interested in hybrid forms of output and ways to evaluate them because these kinds of products are becoming a typical output for humanities research, well aware that academic output could imply societal impact by uptake and use in non-academic settings.

It was decided to work according to a bottom up procedure, giving the field maximum opportunity to influence the approach. All humanities research schools received a questionnaire intended to elucidate characteristics of the research culture of each school asking what channels of communication (journals, publishers) are important for the domain of the school, and what the various academic and other audiences are that they aim at serving. The boards of the research schools appointed panels of prominent members of the schools in order to answer the questionnaire. This resulted in lists of journals and publishers, along with numerous suggestions of other types of research outcomes, such as catalogues, databases, software or documentaries. With regard to questions about hybrid publications the panels identified a number of specific publications that show characteristics of hybrid publications, addressing multidisciplinary audiences in academia and beyond.

While we cannot draw robust conclusions on the basis of the number of publications submitted (we did not ask for a specific number of publications), it is notable that there were differences between fields that might indicate differences in the use of hybrid publications. Eleven panels mentioned in total 156 different hybrid outputs, ranging from five for Archeology to 20 for Gender Studies and 47 for Arts. Hybrid outputs mostly included books and volumes but also catalogues and presentations of exhibitions (Arts), or compendia (Political History).

In both studies the selection of hybrid publications has been performed by researchers working in the respective fields of social sciences and humanities. However, as we shall argue on the basis of a brief systematic analysis of the societal and scientific reception of a selection of these publications, the hybrid characteristics can be demonstrated by showing the attention of stakeholders. For our argument, those publications have been selected that had notable societal impact (as indicated with CRA) and notable numbers of citations as indicated in Google Scholar as a measure of academic impact (Prins et al. 2016).

7 Some of the material presented here is also published in J.B. Spaapen, A. Prins, Contextual evaluation of multi-, inter-, and transdisciplinary research, in: Bernard Hubert et Nicole Mathieu et al. (Eds) Interdisciplinarités entre Natures et Sociétés, Peter Lang, 2016

THE METHOD: CRA

In all cases we present here we gained insight in the variation of output and stakeholders by applying the Contextual Response Analysis (CRA). The analysis is based on an assumption similar to citation analysis, namely that identifiable and unique traces of a publication found on the internet and in specialized databases represent meaningful forms of use, in particular if these traces are linked to identifiable and relevant users. Firstly, traces on the internet and in specialized databases are not arising from publications but from the actions of users that can be seen as a response to the publication even though the nature of this response is unknown. Secondly, by paying close attention to the identification of the users, the response can be placed in context. Identifying users is in part necessary to exclude traces that cannot be regarded as a meaningful response such as those traces that emerge due to fully automated editing of websites. Identifying users also offers the opportunity to place the response into a context of use, such as characterized by the domain in which the response emerges (e.g. News media), or by the characterization of the user on the basis of social function or social-economic sector (e.g. Education, Individuals (Blogger), For-Profit Services etc.). An implication of the method is that the empirical data about use are restricted to traces of users that maintain institutional structures and related infrastructure, in the form of websites and databases. As with any analysis based on internet traces, the societal use by the unconnected population or by those who have little means or time to maintain websites or blogs goes unnoticed. However, the method attempts to focus particularly on the variety of use, comparing both the diversity of use of various products of one institute and the diversity among users among different institutes.

The Contextual Response Analysis is about interest and uptake. It has been developed to shed light on the stakeholder environment by identifying who in the environment of the research group or institute is interested in what is produced by the researchers. To be able to trace the interest of stakeholders and the uptake of articles, reports and other output (films, exhibitions), it is a crucial prerequisite that stakeholders of whatever background have open access to the output of a group or institute.

For this paper, we have focused on one specific form of output: publications. Specific keywords from titles of publications were scrutinised by using search engines such as Google, Yahoo and Bing and also Google Scholar. The results include the complete set of search results of each search engine.[8] The websites, blogs and other (social media) traces that bear references to the publications are identified for various characteristics. These include the social or economic sector in which the referring site operates, such as for example the cultural sector (museums, cultural magazines or blogs devoted to culture), the sector of education, of government, of the health sector or other characterizations that are expected to be relevant for the mission of the investigated research unit. Next, the referring websites, blogs and other traces are also characterized for their function, i.e. their specific communicative role in the sector (does the site belong to a library or repository, a knowledge platform, a book seller, a professional association, a for- profit or non-profit enterprise or an event such as a lecture or conference, or a blog, a publisher etc.). This way the environment can be charted by identifying stakeholders using these publications, and knowledge can be gathered about the uptake of research and the diversity of stakeholders in non-academic as well as academic settings.

In all cases we found a variety of stakeholders related to the output.

This is evident from figures 1 and 2 that show response profiles of the most prominent sectors of stakeholders of the institutes, vis-a-vis a selected number of publications that we analyzed. The selection was done by the board of the institutes on the basis of what they perceived as publications that represent best the societal mission of their institute. The profiles are based on the number of stakeholders that refer to a specific publication. The CRA method excludes self-references as well as references by frequently referring book sellers, publishers and libraries. References in scientific journals are also excluded since indexing of scientific journals is more systematically done by Google Scholar than by Google or Bing, and would lead to confounded results in any comparison with Google Scholar results.

ACADEMIC PUBLICATIONS AS HYBRIDS

In the social sciences and humanities the route for societal use of scientific research is often conceived as taking place via popularizations or via reports that address the articulated demands of policy makers, clients and sponsors. However, the route of popularization implies for many a separation of academic and public debate. As our results also show, such a separation is not a prerequisite in the communication with non-academic users as the academic arguments are also relevant for these audiences.

This comes to the fore in the societal use made of hybrid publications. In the case of AISSR, stakeholders are found both in the societal domain, with sites regarding the higher education sector or other academic professionals (including for profit as well as public services) and in the scientific domain. Table 1 shows five frequently used publications of AISSR combined with the numbers of citations as derived via Google Scholar. Clearly, these are not popularizations but hybrid publications, intended as academic publications that serve both scholarly and societal needs.

8 In other studies in which we applied this method the results of these internet search results have also been compared to those in specialized databases such as Lexis Nexis. In this paper, however, we compare only results of searches on the internet with searches in Google Scholar.

Table 1: Five frequently used AISSR publications used both on internet (Google and Bing) and cited by Google Scholar.

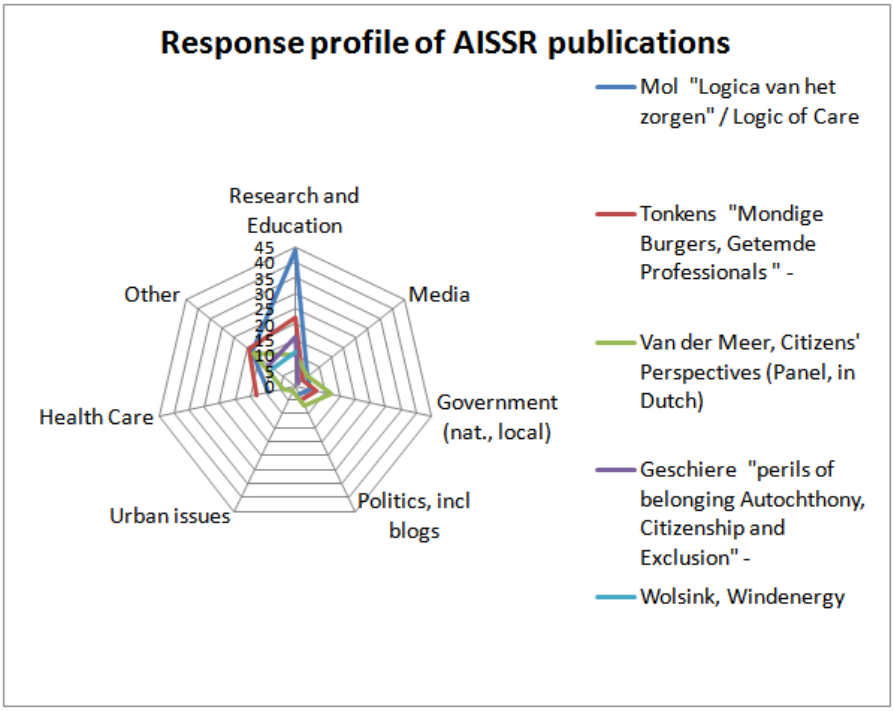
	Kind of publication	Societal use (= Number of unique users)1)	Scientific impact (Citations in Google Scholar)
A. Mol, The Logic of Care, Health and the Problem of Patient Choice, 2008/ Logica van het zorgen, v Gennep 2005 (Dutch version of the book)	Book	48/36	77/532
E. Tonkens, Mondige Burgers, Getemde Professionals, Marktwerking, Vraagsturing en Professionaliteit in de Publieke Sector, NIZW, 2003	Book	69	179
Dekker, Paul, T. Van der Meer, and I. De Goede. „Continu onderzoek burgerperspectieven." Kwartaalbericht 2009 (2009).	Panel	57	16
P. Geschiere, The Perils of Belonging, Autochthony, Citizenship and Exclusion in Africa and Europe, Univ. of Chicago Press, 2009	Book	30	281
M. Wolsink, Wind power implementation, Renewable and Sustainable Energy Reviews, 2007	Scientific Article	27	354

1) Societal use without self-references as well as frequently referring booksellers, publishers, libraries and exclude references by scientific journals

The societal use of hybrid publications also emerges from a more detailed analysis of the stakeholders of the selected products analyzed here. More than one third of the stakeholders are active in the sector of education and research, of which half of them universities or research institutes, often also related to curricula in higher education. We also found stakeholders in a rather variegated context, ranging from democracy projects in Botswana, firms of lawyers in Bangladesh, to bloggers on developments in the National Health Service in the UK, to Dutch local authorities and management consultancies in health care.

In fig. 1 we show the results of a CRA of five hybrid publications of AISSR. Two publications in Dutch were written primarily with the dual purpose of reaching scientific as well as Dutch professional audiences. The book Logica van het Zorgen, the Dutch version of The Logic of Care, is an anthropological and philosophical study of the concept of care in the context of the commercialization of health care practices. It is used by professional organizations delivering or supporting health care, academic hospitals, higher education curricula and health care insurances. The book frequently reaches knowledge platforms on health care, blogs of health care professionals concerned with the developments in health care and also for profit services involved in innovating health care. Also various advocacy groups, including professional organizations and patient groups use this book. The sociological study Mondige Burgers (Expressive Citizens) aims to address the strained relation between vocal citizens and professional expertise, which again leads to use by for-profit organizations with a mission in innovating professional expertise, blogs, and advocacy groups. Two other AISSR studies with a more academic intent (a scholarly book by the anthropologist & sociologist Geschiere, The Perils of Belonging, Autochthony, Citizenship and Exclusion in Africa and Europe, and a paper by the environmental geographer Wolsink, Wind Power Implementation in the scientific journal Renewable and Sustainable Energy Reviews) nevertheless have varied societal stakeholders. Firstly, these publications are used in curricula of universities and in research institutes. Also, advocacy groups are among the more frequent stakeholders. The book about citizenship draws attention from civil activist groups, academic circles outside universities involved in international politics, democracy and citizenship, while the paper about wind energy draws attention from advocacy groups pro and contra wind energy. The volume by Van der Meer is a series with yearly updates about citizens' experiences in relation to government and governance. The volumes are often used by local and regional governments and also by researchers.

Figure 1: Societal use of AISSR publications according to sector, numbers of stakeholders



In a similar way, the hybrid characteristics of the publications in the Humanities can be demonstrated quantitatively. The 127 publications put forward by a number of panels have been investigated initially by using Google as a search engine. As this resulted in a vast number of different websites (over 13.000) a selection of the publications has been investigated further for various characteristics. This allows also excluding various search results that do not reflect the kind of productive interaction specified in the SIAMPI approach. Among these are web shops, booksellers, and also publishers, or libraries and repositories. As references found with generic search engines from scientific journals indicate communication within the field of academia rather than societal communication, these results also have been excluded.

Although each of these publications has drawn attention of societal stakeholders, their reception among academics too is noticeable in citations found with Google Scholar, characterizing the publications as hybrids (table 2).

Table 2: Thirteen frequently used Humanities publications used both on internet and cited by Google Scholar.

	Google Scholar cites	# net societal stakeholders	Panel
Annemarie Mol (2003) The Body Multiple Duke University Press	3359	86	Science and Technology Studies
José van Dijck. The Culture of Connectivity: A Critical History of Social Media. Oxford: Oxford UP, 2013.	729	87	Literature Studies
James C. Kennedy, Nieuw Babylon in aanbouw. Nederland in de jaren zestig (Amsterdam 1995: Boom) (1)	280	132	Political History
Piet de Rooy, Republiek van rivaliteiten. Nederland sinds 1813 (Amsterdam: Metz & Schilt 2002) (2)	151	46	Political History
Ernst van de Wetering. Rembrandt. The Painter at Work, AUP, 1996.	150	106	Arts and Art History
Trudy Dehue (2008) De depressie epidemie, Amsterdam: Augustus	103	206	Science and Technology Studies
Frits van Oostrom, Het woord van eer (1987) & Wereld in woorden (2013)	74	22	Philosophy
Leo Lucassen & Jan Lucassen , Winnaars en verliezers. Een Nuchtere Balans Van Vijfhonderd Jaar Immigratie (Amsterdam: Prometheus, 2011)	65	84	Economic History
M. de Winkel, Fashion and fancy : dress and meaning in Rembrandt's paintings, Amsterdam (AUP) 2006	49	45	Arts and Art History
Henk te Velde, Stijlen van Leiderschap. Persoon en politieke van Thorbecke tot Den Uyl (Amsterdam: Wereldbibliotheek 2002) (3)	41	47	Political History
Marita Mathijsen, De gemaskerde eeuw. Amsterdam, Querido, 2002. 268 pp.	38	57	Cultural History
Floris Cohen, Herschepping van de wereld. Het Ontstaan Van De Moderne Natuurwetenschap Verklaard. Amsterdam, Bert Bakker, 2008. 299 pp.	35	66	Cultural History
Wagenaar, Town planning in the Netherlands since 1800, 2011	25	10	Arts and Art History

From an academic perspective these publications reflect broad academic arguments of historical, philosophical and sociological background, often offering also critical perspectives. In this sense, the reception among academics may extend disciplinary boundaries.

Such broad perspectives extending disciplinary boundaries are widely valued by academic and non-academic audiences alike. This is the case with publications such as Cohen's book about the history of the natural sciences (Herschepping van de Wereld), Mathijsen's study of mentality in nineteenth century Netherlands, or Van Oostrom's history of court culture in The Netherlands (Woord van Eer).

Hybrids are often also topical to existing political and cultural debates, offering information, insights and perspectives on current or recent phenomena, such as Kennedy's study about Dutch culture in the 1960s (Nieuw Babylon), or Te Velde's study about leadership in Dutch politics (Stijlen van Leiderschap), or De Rooij's historical account of how rivalry has shaped Dutch politics throughout the last two centuries (Republiek van Rivaliteiten), or Van Dyck's Culture of Connectivity, a critical history of social media, such as Facebook, Twitter, Youtube and Wikipedia. They received much attention in blogs, curricula at other universities, or knowledge platforms in education. These historical books are not only reviewed in journals and books but also used by politicians, op-ed writers and (local) governments.

Topical is also Mol's book, The Body Multiple, addressing issues with an ethnographic analysis on the various perspectives and practices related to atherosclerosis giving insight in both doctors and patient narratives. The book has been reviewed in non-academic journals, and referred to in blogs (social media), knowledge platforms and by non-academic associations, in a wide variety of sectors.

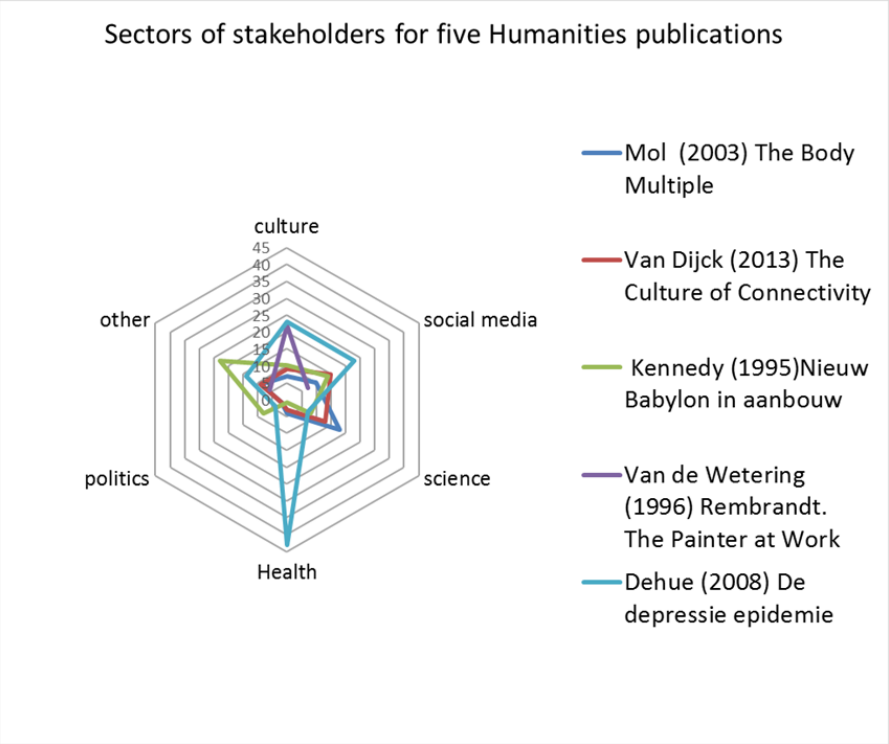
Hybrid work may also imply or involve audiences in specific sectors. This is the case with Van de Wetering's book about working methods of Rembrandt. Since this book addresses also issues about authenticity, users in the cultural sector are keen to use and also discuss or debate this work because of the implications of assigning particular paintings or drawings to Rembrandt. Similar professional interest can be noted for the book on Town Planning by Wagenaar, with users in architecture and urban planning.

The topical nature of many hybrids can also lead to fierce debate. This is the case with Lucassen's book offering many facts and perspectives on the heavily contested issue of immigration, with references from bloggers, opinion makers in weeklies, politicians and interest groups working for refugees. Another example of a hybrid sparking fierce public debate is the historical and philosophical study about the rise of the diagnosis of depression by Dehue (Depressie Epidemie)[9]. The study received considerable media attention and also the Dutch award for best science communication book in 2009 by the Dutch Science Foundation NWO. The award praised the book to present "work that is relevant both scientifically and in society". The laudation read that "with historical analysis and well-funded methodology and analysis in science theory […] and genuine apprehension about recent developments unmasks as unreliable ice what has been held for solid foundation". However, academic psychiatrists were less taken away by the book, criticizing it as "a poisonous broth that is harmful … one is almost to compare it with Fitna" (a controversial film by the Dutch populist politician Geert Wilders).[10] The book led to many appearances

9 Disclosure: Prins is the partner of Dehue. The selection of the book as a hybrid, however, was done by the panel of the school for Science, Culture and Technology Studies.

on tv-shows and radio, interviews in and contributions to newspapers and weekly's, lectures for professional and non-professional audiences, and invitations for parliamentary special committees on health.

Figure 2 Societal profile of publications in Humanities



STAKEHOLDER POLICIES: ACTIVELY MIXED PORTFOLIOS

These variegated patterns of use show that the contexts in which institutes and researchers work are multi-faceted. Also, stakeholders – researchers, professionals, policy makers, and citizens- are not operating in neutral environments or circumstances but are engaged in wider professional as well as political debates.

The awareness of researchers and their institutes for the contextual complexity becomes apparent in how the interactions with the environment are managed. Elsewhere, we have shown that policy research institutes outside academia may use more or less formal ways to organize interactions with environments (Spaapen and Prins 2016, Spaapen and Van Drooge, 2011, see also Hage et al 2010). However, in contrast to the explicitly organized stakeholder relations by policy research institutes, academic settings such as AISSR or schools in the Humanities appear to represent a distinct modality. Firstly, stakeholder connections are loosely organized by individual researchers or research groups. Secondly, stakeholder relations are maintained while at the same time fulfilling an academic mission. While this does not prohibit stakeholder connections, it does complicate the evaluation of researchers' output. As institutes are competing with other institutes for funds and material support, within the university but also outside, they are always under pressure to perform evaluations that do not divert too much from what is prevailing. This tension shows for example in the mission statement of AISSR, where it reads that there is ample room for researchers to invest in creative and unusual approaches, rather than holding to a demand-driven research agenda. But at the same time, it is the policy of this institute to allow for a wide diversity of publication types, including monographs, contributions in newspapers as well as articles in journals. This enables groups and individual researchers to combine various publication channels.

We see something similar with the output in the Humanities research schools. The variegated output of these schools targets a wide variety of audiences, both within and outside academia. And, as the examples in table 2 show, there is a broad societal interest in the hybrid publications of these schools. And the societal uptake of each of these books is supported by the authors with public appearances in the form of lectures for various audiences, op-eds in newspapers, essay contributions in magazines or media appearances on radio, tv or in documentaries, blogs etc. Evidence for sure, that the context in which researchers operate is changing and that this influences their modus operandi.

While the hybrid publications noted above represent original scientific work, their critical characteristics makes them accessible for wider audiences and enable debates with relevant stakeholders. They also may help stakeholders to reflect on their professional practices. Apparently, the effectiveness of the stakeholder connections are not frustrated by the fact that researchers also claim room for 'pure' academic endeavors, nor are these latter held up by the authors engaging in societal debate via indirect forms of interactions such as newspaper columns combined with direct interactions with stakeholders through lectures, e-mail contacts or participation in committees. Clearly, whether intended or not, authors of hybrid work maintain a portfolio of different types of communication with the variety of stakeholders. Such portfolios are tailored for the specific sectors of the different stakeholders, depending for instance on the social characteristics of the sector, such as the degree of institutionalization or occurrence of larger organizations (political parties, governments, larger museums) as opposed to loose sets of individual users or small organizations such as artists, writers, general practitioners or health care workers). And it is time that this colorful palette gets recognized in regular evaluation procedures.

10 The quotes are taken from the website of the author http://www.trudydehue.nl/boeken.html and translated from Dutch.

IMPLICATIONS OF DIVERSITY OF STAKEHOLDERS AND MIXED PORTFOLIOS FOR RESEARCH EVALUATION

Evaluating academic research for both quality and relevance is challenging for everybody involved: researchers, stakeholders, evaluators and policy makers. Researchers have to find a balance in addressing the relevant audiences within and outside academia; while stakeholders have to engage early on in debates in order to be able to participate in shaping common research agenda's, evaluators have to broaden the way they assess the broader output of groups and institutes. And policy makers have to allow for more versatile evaluation procedures in which the context of research is somehow embedded. We have shown that there can be a rather large diversity among interested stakeholders, depending not only on the (multi)disciplinary background of the research but also on the very topic of research. This is apparent in both the environment of the Humanities research schools and that of AISSR, showing distinctly different users such as bloggers and book reviewers for historical books, or professionals in the relevant sectors for works on urban issues, politics or health.

Also, the variegated ways in which communication takes place via the mixed portfolios of researchers shows, that communication is multifaceted, depending on aspects such as topicality in general or specialized debates and the communicative characteristics of the fields in which stakeholders are working and are organized.

The current practices in evaluation often aim at the ranking and benchmarking of research units. This assumes also the possibility of comparison. However, as the examples show, the diversity of the stakeholders, and the multifaceted characteristics of the communication resist such comparison.

The CRA method we presented above serves two purposes at least. It aims at showing the uptake of research results in the societal context of research groups. And, it intends to help researchers to write up their societal impact in a more convincing way. This hopefully helps them in evaluation procedures that allow for a wider perspective on research than is usually the case in traditional metrics oriented evaluations that primarily are looking at output in the scientific literature. Those kinds of evaluations rest on a competitive view on science. But science is not a game, it is a serious business and society that is investing public money in research may expect that researchers are prepared to share and debate their work with interested audiences also outside their own field. Evaluation procedures in many countries start to recognize that as shown by the REF UK 2014 and the Dutch SEP 2015-2021, by allowing room in their protocols for assessment of research impact on society. In the Dutch case, this goes so far that there is a complete balance when it comes to weighing the scientific and the societal impact of research. CRA and other methods like the impact case studies used in the REF are meant to help all involved to gain more insight in the uptake and impact of research in society. Methods like CRA provide information that is not only systematic and robust but enables also to address issues of context and ambition of the research. In this sense, it is possible not only to look back to past performance but to use evaluations as formative events: as assessments of future opportunities and challenges.

REFERENCES

Bornmann, Lutz, (2013), What Is Societal Impact of Research and How Can It Be Assessed? A Literature Survey, Journal of the American Society for Information Science and Technology, 64(2):217–233, 2013

Federation 2017, Federation for the humanities and social sciences Canada, Approaches to Assessing Impacts in the Humanities and Social Sciences, May 2017

Hage, M., Leroy, P., Petersen, A.P., Stakeholder participation in environmental knowledge production. Futures 42.3 (2010): 254-264

Manville, Catriona et al., (2015), Assessing impact submissions for REF 2014: an evaluation Published by the RAND Corporation, Santa Monica, Calif., and Cambridge, UK

Prins, A. A., Costas, R., van Leeuwen, T. N., & Wouters, P. F. (2016). Using Google Scholar in research evaluation of humanities and social science programs: A comparison with Web of Science data. Research Evaluation, rvv049.

Spaapen, J.B. and A.A.M. Prins, Final Report SIAMPI WP5 Health, with two case studies for NIVEL and LUMC, June 2010

Prins, A.A. M., Gebruik en gebruikers van WODC publicaties, Een bibliografische analyse van het WODC 2008-2010, Groningen 2012;

Prins, A.A.M., Contextual Response Analysis of reports of the Netherlands Environmental Assessment Agency PBL, Groningen 2012

Prins, A.A.M., J.B. Spaapen, Contextuele Respons Analyse van publicaties van het Sociaal en Cultureel Planbureau 2010-2015, Groningen 2016

Prins, A.A.M. Contextual Response Analysis of publications of the CPB Netherlands Bureau for Economic Policy Analysis, Groningen, 2015

Spaapen, Jack and Leonie van Drooge, Introducing 'productive interactions' in social impact assessment, (2011), Research Evaluation, 20(3), September 2011: 211–218

Spaapen, J.B. and A.A.M. Prins, Contextual evaluation of multi-, inter-, and transdisciplinary research, in: Bernard Hubert et Nicole Mathieu et al. (Eds) Interdisciplinarités entre Natures et Sociétés, Peter Lang, 2016

Wetenschapsvisie 2025, Keuzes voor de toekomst (Science Vision 2025, choices for the future), white paper of the Dutch government, published in november 2014, The Hague

AUTHORS

A.A.M. PRINS
Support in ResearchManagement
E: info@adprins.nl

J.B. SPAAPEN
KNAW (the Netherlands)
E: jack.spaapen@knaw.nl

NEW INDICATORS FOR OPEN SCIENCE

POSSIBLE WAYS OF MEASURING THE UPTAKE AND IMPACT OF OPEN SCIENCE

DIETMAR LAMPERT, MARTINA LINDORFER, ERICH PREM, JÖRG IRRAN AND FERMÍN SERRANO SANZ

KEYWORDS: open science, uptake and impact indicators, altmetrics, new open science indicators, open science evaluation

ABSTRACT

Open science (OS) opens up new ways of creating and sharing knowledge and of disseminating various kinds of results, such as traditional articles, research data, computational and mathematical codes, 3D models, interactive visualisations, or micro-insights. Moreover, OS offers the chance to introduce new ways of evaluating science in a more nuanced, fair, and precise way. As the recent 'altmetrics' push has shown, there is wide agreement that conventional approaches to science evaluation are inadequate. With the *open* movement becoming stronger – especially in science –, it is a good time to reflect on potential new indicators to gauge the uptake and impact of OS. This conceptual work aims to offer a vantage point for more substantial discussions among the key stakeholders.

INTRODUCTION

The traditional way of evaluating science comprises, among other things, an approximation of impact – typically the number of received citations (cf. Mingers & Leydesdorff 2015; Garfield, Eugene, and Alfred Welljams-Dorof 1992, Weingart 2005). Although the advent of the internet made it significantly easier to calculate such indicators, warning signs of their misuse appeared early (cf. e.g. Kostoff 1998, Gläser et al. 2002, Butler 2003, or Weingart 2005). The critique – not just of the misuse and unintended bad consequences of indicators but also policies and practices that had adverse systemic effects on scientific research – culminated in a series of declarations and manifestos[1] that went hand in hand with the widening of the open movement.

This article does not intent do examine OS (Open Science) from a history point of view (for that, cf. David 1998, 2003; Willinsky 2005; Bartling & Friesike 2014a) nor does it examine the various concepts of OS (for that, cf. Bartling & Friesike 2014b; Fecher & Friesike 2014; Buschmann et al. 2015; Delfanti & Pitrelli 2015). It suffices to recognise that Open science (OS) does not only open up new ways of creating and sharing knowledge, of disseminating results of individual components along the scientific research process but of evaluating science more nuanced, precise, and fair. This means that much is expected of OS, which is further underlined by a recent report by the OECD (2015a) that states that the positive factors associated with OS are, for instance, increasing transparency and quality in the research validation process, improving efficiency in science, or increasing the knowledge spill-overs to the economy.

This is not only a matter of technological developments but also of change in cultural practices. It is yet unclear how the uptake and impact of OS practice ought to be monitored and measured, on research in general but on society in particular. This article is based on the results of a study[2] on Open Science conducted for the European Commission and represents conceptual work as, so far, no substantial work has been done before in this regard.

In the literature review and in the interviews with OS experts that we conducted, there is a general consent that possible new indicators for the monitoring and assessment of scientific production and its impact need to be agreed on by all stakeholder groups, in light of a major redesign of the scientific process provoked by OS.

WHAT IS ALREADY BEING MEASURED

The most prominent attempt to move beyond the traditional impact indicators and towards more open, extensive ones is *altmetrics* (cf. Priem, Piwowar, Hemminger 2012; Galloway & Pease 2013; Bornmann 2014). Although it employs indicators that are enabled by new technology and extend their reach to capture impact on society, the concept is still in its infancy. Moreover, it is yet unclear what altmetrics can actually signify (cf. Mingers & Leydesdorff 2015).

It has become evident (cf. EC 2015), that new evaluation systems are needed – evaluation of research that is not solely based on bibliometric indicators and that does take into account the whole array of contributions to and resulting from the research process (data, methods, code, insights, ideas, trainings, participations in all kinds of activities, etc.). One can see that altmetrics do not go far enough – the open concepts involved in OS exceed that scope by far.

1 see e.g. the Budapest Open Access Initiative (2002), the Berlin declaration on Open Access (2003), the Declaration on Access to Research Data from Public Funding (2004), the Open Science Rome Declaration (2012), the San Francisco Declaration on Research Assessment (2013), the Liber Statement on Open Science (2014), the Leiden Manifesto for research metrics (2015), or even the Amsterdam Call for Action on Open Science (2016)

2 tender SMART 2014/0007 "Open Digital Science"

WHAT TO ACTUALLY MEASURE

One of the main objectives of the study underlying this article was to propose a framework for an OS observatory which monitors the progress of OS in Europe on a continuous basis. The indicators suggested in the article shall therefore be useful to monitor the uptake and impact of OS. Also, indicators shall measure if OS practices make science more accessible for a wider audience, whereby Fecher and Friesike (2014:19) see accessibility in the double sense: (a) accessibility of the research process and (b) comprehensibility of the research result. This understanding suggests that the relationship between science and society must be reflected in the indicators in any case.

Unbeknown to the project, RAND Europe had been tasked by the European Commission to develop the Open Science Monitor that was to accommodate a whole range of indicators to monitor and measure OS trends in the EU. They conducted their work in parallel to our project. Before the writing of this article, we had a chance to scrutinise their results (Smith et al. 2016), which yielded similarities but also differences compared to our results, which we will mention below in the indicators sections.

METHODOLOGY

To come up with reasonably sound results, our project employed a mix of methods that started with a thorough desk research on the status quo of OS concepts, metrics, good practices, policies, programmes, and stakeholders, predominantly in the EU. To better understand the technology characteristics inherited by OS and to predict its potential evolution in the near future from a technological point of view, a trend analysis was conducted.

The next phase consisted of a series of consultations with roughly 60 EU experts from research, industry, policy, and RTD management that was kicked off with interviews on the OS vision, metrics of OS uptake and impact, and the involved main players and surfaced good practices. Based on this work, six distinct future OS scenarios were created to provide the necessary level of concreteness for the development of a first set of OS uptake and impact indicators that were scrutinised through a wider online consultation. Finally, a focus group served to validate the results and explore concrete policy options.

AN INITIAL SET OF NEW INDICATORS

The application of the above-mentioned methodology yielded, among others, a first set of new possible indicators for measuring the uptake and impact of OS. We could observe mainly two major dimensions – one pertains to the scientific process itself, i.e. the way science is conducted; the other pertains to the system level and thus the framework conditions. Each of these two major dimensions has several sub-dimensions:

1. the scientific process:
    - conceptualisation and data gathering/creation
    - analysis
    - diffusion of results
    - review and evaluation
2. the system level:
    - reputation system, recognition of contributions, trust
    - open science skills and awareness
    - science with society

These sub-dimensions are not exhaustive; they merely pose a categorisation that aligns well with the identified, new potential OS indicators. It goes without saying that this categorisation will need to be revised and refined the further the indicators are being developed.

Each of the above-mentioned dimensions entails a cluster of indicators. Those will be presented below in terms of their nature, their relevance, and the stakeholder group responsible for adopting and further developing an indicator. This article will not cover the entirety of indicators elaborated by the project team but only a subset of those sub-dimensions that are most relevant for the theme of the Open Evaluation conference. That said, the other sub-dimensions will at least provide a rough description to provide context and make it easier to understand the scope of the cluster.

Comparing these results with the ones generated by RAND Europe (Smith et al. 2016), there are similarities in terms of indicators that pertain to the scientific process, like open access publications (e.g. percentage of publications from each year that are open access, rate of green open access publications compared to journal publications, number of preprints, or journal policies on open access), open research data (e.g. number of data repositories, or funder policies on data sharing), and open scholarly communication (e.g. percentage of peer reviews that are published, journal policies on open peer review, use of altmetrics platforms/number of mentions of publications in media and social media, or articles published before peer review). Their work offers little with regard to the system level, though, which is the biggest difference compared to our work. Our consultation has shown that the necessary framework conditions need to be in place to foster an *open* culture.

Figure 1: Stakeholder groups - abbreviations and colour

R	researchers
RO	research (conducting) organisations
RFO	Research-funding organisations
PM	policy-makers
PU	publishers

The presented indicators contain the stakeholder group that is – not solely but – mainly responsible for further developing and adapting an indicator. In some instances, more than one stakeholder group is responsible, i.e. when an indicator is fairly complex to design, maintain, or yield data. In any case, these stakeholder groups are defined as follows:

Each presented indicator will also have a mean rating that pertains to the consulted experts' view on the relevance of said indicators – a 10 means the highest relevance, 0 no relevance at all; we have eliminated all of the roughly 60 indicators that did not achieve an above-average rating of at least 7.5.

INDICATORS CLUSTER I: CONCEPTUALISATION & DATA GATHERING/CREATION

Important questions in this dimension are whether the quality of data and information is adequate, e.g. whether the data were properly cleaned, whether they are curated, whether metadata are provided, etc. Recent policy trends involve mandatory rules and requirements (most commonly, funding agencies mandate public access to funded research), and the development of infrastructure to enable OS. Fewer initiatives relate to non-monetary incentive mechanisms like the definition of new reward/promotion systems.

Scientific work must no longer be restricted to measuring final products (such as articles), but should measure the development of the individual steps of the scientific workflow. Furthermore, results will differ according to disciplines, fields, or data types. Indicators in this dimension cover e.g. research funding organisations requiring the open provision of data/code, the accessibility of data/code, or the availability of metadata.

Requirements from research funders	mean rating (0..10 max.)
% of research funders that mandate the provision of the data / software code produced in the context of the funded activity AND who mandate the conformity to data (exchange) standards	7.9
	RFO PM

Accessibility	mean rating (0..10 max.)
accessibility of open data / code as % of all data / code produced by publicly (co-)funded projects	9.1
	R RO RFO

Machine-readable	mean rating (0..10 max.)
% of machine-readable data / metadata	7.9
	PU R RFO

Availability of metadata	mean rating (0..10 max.)
availability of explanatory metadata as % of all available data (resulting from publicly (co-)funded research)	7.5
	PU R RFO

Quality of metadata	mean rating (0..10 max.)
quality of metadata (versioning, volume, data format, description of fields, etc.)	8.2
	PU R RFO

Simulation results	mean rating (0..10 max.)
usability of simulation results (models, data, and code)	7.5
	R RFO PU

Data services	mean rating (0..10 max.)
(types of) open data services offered	8
	PU R RO

Data compilation/publication costs incorporated	mean rating (0..10 max.)
% of funded projects incorporating costs for data compilation / publication and maintenance (of the repository/data sets)	7.6
	PM RFO RO

Long-term availability	mean rating (0..10 max.)
is the (long-term) availability of the data guaranteed (availability of a sustainability plan (yes/no))	8.2
	RFO RO PM

Sharing policies	mean rating (0..10 max.)
# of sharing policies in research organisations (sharing of data, organisms, etc.)	7.6
	RO

INDICATORS CLUSTER II: ANALYSIS

Respondents in this cluster argue that open methods contribute to improving the reliability of research results but that the impact of the open methods were still marginal because their use is not spread widely yet in the research community. Indicators in this cluster that are easier to design and monitor are data citations[3] and code/software citations, a possible new one might be content citations.

INDICATORS CLUSTER III: DIFFUSION

We deliberately chose the term "diffusion" (of results) instead of the term "publication" which is most commonly used in academia". We want to stress that diffusion can and – some would argue – should start well before the results are out. In our online assessment, several comments underpinned the need to get away from the traditional paper publishing models and find indicators that gauge the growth of dissemination channels other than journals. Participants stated that journals are becoming irrelevant in many fields already. Impact of OS can more easily be captured in those cases where open communication and responsive attitude to feedback have actually changed the trajectory of research, e.g. a side-line turned into the main thing, a bug/design issue was detected, or the project just responded (or even emerged in response) to what is happening in society.

INDICATORS CLUSTER IV: REVIEW AND EVALUATION

Currently, peer review is the standard practice to assure quality of scientific output. Traditional peer review has well known shortcomings, though, such as little credit given to reviewers, lack of transparency and limited verification of scientific results (cf. OECD 2015). Open peer review is often mentioned as an alternative, but not without the same amount of criticism. In the Open Science community, however, there is certain agreement that transparency measures need to be taken in the review and evaluation process. A multitude of suggestions have been put forward, some of which are considered as "incremental", meaning that they would not do much harm to the current review procedure, while others as regarded "radical" or transformative. Adding transparency to the review process can happen at various stages. One option would be to make grant proposals publicly accessible at various points of time, e.g. after the project has ended, along with the final project reports, at the beginning of a project, at the point of announcing funding decisions, upon submission to the funder and during the drafting phase (cf. Mietchen 2014). Another would be to make the peer review public. This can again happen in an incremental form, meaning that some knowledge within the peer review process is made openly accessible, or in a radical form, meaning that transparency of knowledge becomes a separate pillar of legitimacy itself (cf. Gurwitz, Milanesi, and Koenig 2014). Open peer review is currently a highly contested field and so is the choice of respective indicators. This can also be said for the question how societal relevance of research should be treated and assessed in evaluation. A rather easy measure could be to make the "impact statement" of a proposal publicly accessible. A labelling system for expected impact (oriented on e.g. the Sustainable Development Goals) could be an option to create clearer evaluation references. Again, there are several options to develop new indicators but only a few concrete ones passed the threshold or were further suggested.

Openness in calls for proposals	mean rating (0..10 max.)
openness in call for proposals (open proposals, open submissions, open review)	7.8
	PM RFO RO

Review criteria	mean rating (0..10 max.)
% of peer reviews that include reproducibility and transparency as review criteria	7.7
	RFO PU

INDICATORS CLUSTER V: REPUTATION SYSTEM, RECOGNITION OF CONTRIBUTIONS, TRUST

The uptake of OS practice in the research process is unlikely to flourish if researchers fear it is not properly acknowledged and officially recognised. This is underpinned in the initially mentioned surveys on researchers' attitudes towards OS, which reveal low factual progress in putting OS into practice. Reward mechanisms for data sharing are currently especially weak and researchers might choose rather not to spend a serious amount of time in cleaning and curating their data for the reuse of others. Some organisations (datacite, ORCID, Figshare, Dryad Digital Repository, ResearcherID) have propositions for data citation tools which would credit authors for data and metadata sharing, but "in most countries the existing framework does not promote sharing efforts, especially with respect to results, data sets or other research material at the pre-publishing phase" (OECD 2015a, p. 89). Formal recognition of a variety of contributions along the scientific process (e.g. to the selection of research topics, formulation of hypotheses, project participations, review activities, etc.) has yet to be adopted. To understand the importance of the recognition of contributions, it serves to recall the various roles that are involved in the scientific process (see figure below).

3 Platforms that may provide data on data citation: DataCite, ORCID , Figshare, The Dryad Digital Repository, ReseacherID.

Figure 2: Roles in the scientific process. Source: Liz Allen et al. (2014): Credit where credit is due; Amy Brand, Liz Allen, Micah Altman et al. (2015): Beyond authorship: attribution, contribution, collaboration, and credit.

Term	Definition
Conceptualization	Ideas; formulation or evolution of overarching research goals and aims
Methodology	Development or design of methodology; creation of models
Software	Programming, software development; designing computer programs; implementation of the computer code and supporting algorithms; testing of existing code components
Validation	Verification, whether as a part of the activity or separate, of the overall replication/reproducibility of results/experiments and other research outputs
Formal Analysis	Application of statistical, mathematical, computational, or other formal techniques to analyze or synthesize study data
Investigation	Conducting a research and investigation process, specifically performing the experiments, or data/evidence collection
Resources	Provision of study materials, reagents, materials, patients, laboratory samples, animals, instrumentation, computing resources, or other analysis tools
Data curation	Management activities to annotate (produce metadata), scrub data and maintain research data (including software code, where it is necessary for interpreting the data itself) for initial use and later reuse
Writing – Original Draft	Preparation, creation and/or presentation of the published work, specifically writing the initial draft (including substantive translation)
Writing – Review & Editing	Preparation, creation and/or presentation of the published work by those from the original research group, specifically critical review, commentary or revision – including pre- or post-publication stages
Visualization	Preparation, creation and/or presentation of the published work, specifically visualization/data presentation
Supervision	Oversight and leadership responsibility for the research activity planning and execution, including mentorship external to the core team
Project Administration	Management and coordination responsibility for the research activity planning and execution
Funding acquisition	Acquisition of the financial support for the project leading to this publication.

Although the importance of this sub-dimension has been recognised, only one of the suggested indicators was rated high enough to reach the predefined threshold.

Data communication as valued scientific contribution	mean rating (0..10 max.)		
data communication recognised as criterion for career progression (yes/no)	7.5		
	RO	**R**	**PM**

Further options to explore are for example the % of publications in Open Access Journals (with or without impact factor) or availability of means to easily publish negative results.

INDICATORS CLUSTER VI: OS SKILLS & AWARENESS

OS-related skill development across disciplines will be a crucial factor for the maturation of OS in Europe. Researcher's skills in OS (e.g. curating and maintaining large data sets) differ across disciplines due to different traditions or training opportunities in digital tools and data handling. There is a substantial need for further training of researchers and scientists in handling big, multi-layered and complex data sets. Accordingly, indicators in this cluster cover e.g. the monitoring of skilled personnel, research personnel active in OS, or the awareness and use of open standards.

INDICATORS CLUSTER VII: SCIENCE WITH SOCIETY

This cluster is about finding indicators that assess effects of OS on the promotion of the engagement of citizens in science and research. As Mietchen, Mounce, and Penev (2015) observed, most of the research process is hidden from public view through multiple layers of obfuscation that stems from inherited conventions and habits from the paper era. This has begun to change, though, not least because digital technologies enable engagement and popularisation. Popularisation activities are understood as targeting a wide audience and a non-specialised public. Consequently, relevant new indicators gauge, among others, citizens' engagement in (open) science, research communication (beyond academia), or the accessibility of data that are of public interest.

CONCLUSION AND OUTLOOK

Designing indicators to measure the uptake and impact of OS (Open Science) is a challenge, not least because the concept itself is still evolving. OS is necessarily broad because it is composed of many dimensions (e.g. along the scientific research process) and embedded in a larger system that involves e.g. new skills, a new reputation scheme, or the wider public.

Most indicators proposed in this article are new and not gathered/surveyed/evaluated automatically (yet). Consequently, a first vital step is to put the necessary mechanisms in place. To achieve this, we propose stakeholder groups that are primarily involved in/responsible for designing, measuring, interpreting, and/or adapting an indicator.

It should be of prime concern to avoid the early mistakes of bibliometrics that had severe unintended negative consequences on the research system. An essential precondition to circumnavigate Campbell's law and to make indicators work as intended is that all concerned stakeholder groups are involved in their design and evolvement. They all need to agree on what an indicator should measure (and what it should not) and how it should be used (and what it must not be used for). Furthermore, indicators need to be flexible enough to accommodate differences, e.g. in research fields, and allow the emergence of new developments. The differences in research fields can be considerable, as is the pace at which OS is being adopted in those fields. Those differences will need to be elaborated and reflected in the respective indicators.

Furthermore, all stakeholders need to make sure that the OS indicators are and remain a means to an end and never become an end in themselves; otherwise, Campbell's law would apply again.

Finally, new indicators need to be tested – not just discussed – before being adopted on a larger scale. This can be done in small experiments by using individual, selected indicators.

BIBLIOGRAPHIC REFERENCES

Liz Allen et al. (2014): Credit where credit is due.

Amy Brand, Liz Allen, Micah Altman et al. (2015): Beyond authorship: attribution, contribution, collaboration, and credit.

Alperin J.P., Babini, D., Fischman, G.E. (Eds., 2014): Open access indicators and scholarly communications in Latin America. Buenos Aires: CLACSO, ISBN 978-987-722-027-8. CLACSO virtual library.

Bartling, Sönke, and Sascha Friesike (2014a): Towards another scientific revolution. Opening science. Springer International Publishing, 2014. 3-15.

Bartling, Sönke, and Sascha Friesike (2014b): Opening science: The evolving guide on how the internet is changing research, collaboration and scholarly publishing. Springer Open.

Bornmann, Lutz (2014): Measuring the broader impact of research: The potential of altmetrics. No. arXiv: 1406.7091.

Buschmann, Katrin, Stefan Kasberger, Peter Kraker, Katja Mayer, Falk Reckling, Katharina Rieck, Michela Vignoli (2015): Open Science in Österreich: Ansätze und Status (Open Science in Austria: Approaches and status), in Information. Wissenschaft & Praxis 2015; 66(2–3): 1–9. DOI 10.1515/iwp-2015-0025

Butler, Linda (2003): Modifying publication practices in response to funding formulas. In Research Evaluation (2003), 17

(1) : 39–46; DOI: https://doi.org/10.3152/147154403781776780

David, Paul A. (1998): Reputation and Agency in the Historical Emergence of the Institutions of 'Open Science'. Centre for Economic Policy Research, Publication No. 261, Stanford University (March 1994, rev. Dec. 1998)

David, Paul A. (2003): The Economic Logic of "Open Science" and the Balance between Private Property Rights and the Public Domain in Scientific Data and Information: A Primer. In The Role of the Public Domain in Scientific and Technical Data and Information , Washington, DC: National Academies Press, 2003

Delfanti, A., Pitrelli, N. (2015): Open science: revolution or continuity? In: Open Science, Open Issues, Series UC Davis Previously Published Works

EC (2015): Indicators for promoting and monitoring Responsible Research and Innovation. Report from the Expert Group on Policy Indicators for Responsible Research and Innovation, EUR 26866 EN

Fecher, Benedikt, and Sascha Friesike (2014) Open Science: One Term, Five Schools of Thought. Opening Science: 17.

Finnish Ministry of Education and Culture (2014): Open science and research leads to surprising discoveries and creative insights. Open science and research roadmap 2014-2017. Reports of the Ministry of Education and Culture, Finland 2014:21

Galloway, Linda M. and Pease, Janet L. (2013): Altmetrics for the Information Professional: A Primer. Libraries' and Librarians' Publications. Paper 105.

Garfield, Eugene, and Alfred Welljams-Dorof (1992): Citation data: their use as quantitative indicators for science and technology evaluation and policy-making. Science and Public Policy 19.5 (1992): 321-327.

Gläser, Jochen, Grit Laudel, Sybille Hinze, and Linda Butler (2002): Impact of evaluation-based funding on the production of scientific knowledge: What to worry about, and how to find out. Expertise for the German Ministry for Education and Research 31 (2002).

Jensen, P., Jean-Babstiste, R., Kreimer, P., Croissant, Y. (2008): Scientists who engage with society perform better academically, http://arxiv.org/abs/0810.4672

Kostoff, R. N. (1998): The use and misuse of citation analysis in research evaluation. In Scientometrics, Vol. 43, No. 1 (1998) 27-43

Mietchen D, Mounce R, Penev L (2015): Publishing the research process. Research Ideas and Outcomes 1: e7547. doi: 10.3897/rio.1.e7547

Mietchen, D. (2014): The Transformative Nature of Transparency in Research Funding. Published: December 30, DOI: 10.1371/journal.pbio.1002027

Gurwitz, David, Elena Milanesi, and Thomas Koenig (2014): Grant application review: The case of transparency. PLoS biology 12.12 (2014): e1002010. https://doi.org/10.1371/journal.pbio.1002010

Mingers, John and Loet Leydesdorff (2015): A Review of Theory and Practice in Scientometrics. European Journal of Operational Research

OECD (2015a): Making Open Science a Reality. OECD Science, Technology and Industry Policy Papers, No. 25, OECD Publishing, Paris. http://dx.doi.org/10.1787/5jrs2f963zs1-en

OECD (2015b): OECD Digital Economy Outlook 2015. OECD Publishing, Paris. DOI: dx.doi.org/10.1787/9789264232440-en

Prem, Erich, Fermín Serrano Sanz, Martina Lindorfer, Dietmar Lampert, Jörg Irran (2016). Open Digital Science. Technical report.

Priem, Jason, Heather A. Piwowar, Bradley M. Hemminger (2012): Altmetrics in the Wild: Using Social Media to Explore Scholarly Impact. https://arxiv.org/html/1203.4745

Scheliga, K., Friesike, S. (2014): Putting Open Science into practice: A social dilemma? http://firstmonday.org/ojs/index.php/fm/article/view/5381/4110

Smith, Elta, Salil Gunashekar, Catherine Lichten, Sarah Parks, and Joanna Chataway (2016): Informing science and innovation policies: Towards the next generation of data and indicators. OECD Blue Sky III Forum, Ghent (BE), Sep. 2016. https://www.oecd.org/sti/063 - OECD Blue Sky 2016_Open Science.pdf

Weingart, Peter (2005): Impact of bibliometrics upon the science system: Inadvertent consequences? In Scientometrics, Vol. 62, No. 1 (2005) 117–131

Willinsky, John (2005): The unacknowledged convergence of open source, open access, and open science. First Monday 10.8 (2005).

AUTHORS

corresponding author:
DIETMAR LAMPERT
ZSI – Centre for Social Innovation, Austria
Linke Wienzeile 246, 1150 Vienna
T: +43 1 4950 442
E: lampert@zsi.at

MARTINA LINDORFER
ZSI – Centre for Social Innovation, Austria

ERICH PREM
eutema GmbH, Austrian

JÖRG IRRAN
eutema GmbH, Austrian

FERMÍN SERRANO SANZ
Universidad de Zaragoza, Spain

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
W carotte.at
E caroline@carotte.at

PRINT
REMA-Print-Littura Druck- u. Verlags Ges.m.b.H
Neulerchenfelderstraße 35, A-1160 Vienna

AUSTRIAN PLATFORM FOR RESEARCH AND TECHNOLOGY POLICY EVALUATION
Platform fteval members:

Federal Ministry of Science, Research and Economy (bmwfw), Federal Ministry of Transport, Innovation and Technology (bmvit), Austrian Institute of Technology (AIT), AQ Austria, Austria Wirtschaftsservice (AWS), Christian Doppler Research Association (CDG), convelop cooperative knowledge, Austrian Research Promotion Agency (FFG), Austrian Science Fund (FWF), Institute for Advanced Studies (IHS), Austrian Institute for SME Research (KMFA), Joanneum Research, Ludwig Boltzmann Society (LBG), Austrian Institute of Economic Research (WIFO), Austrian Council for Research and Technology Development, Austrian Academy of Science (ÖAW), Technopolis Austria, Vienna Science and Technology Fund (WWTF), Wirtschaftsagentur Wien. Ein Fonds der Stadt Wien, WPZ Research, 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