EUROPEAN SCIENCE FOUNDATION

Research Funders and Research Output Collection

Working Group3: Comparative Research Portfolios

ESF Member Organisation Forum on Evaluation of Publicly Funded Research

Executive Summary

The European Science Foundation (ESF) Member Organisation Forum on Evaluation of Publicly Funded Research has examined current practices and plans concerning the collection and analysis of output data. The initial stage of this work has been published previously[1]. This report completes this work by examining the challenges in defining, collecting and analysing information about publications, intellectual property, products and interventions, policy influence, training, collaboration, dissemination activities and use of research infrastructures. This is not an exhaustive list of research outputs, but the discussion highlights the need for output collection to be carefully designed and quality controlled.

The definition, collection, analysis and sharing of output data is a fast moving field. Only three to four years ago there were few systematic, and no cross-funder, approaches to gathering information about research outputs. In 2011/12 the first cross-funder systems for output collection were launched in the UK; in the next few years there may be five to ten systematic large-scale output collection initiatives across Europe alone.

This increased interest in gathering this information makes addressing the challenges in linking, validating and achieving best value from these data ever more important. Research organisations will want to move from a focus on local evaluation of progress, productivity and quality to national and international analysis of impact and more sophisticated benchmarking and comparisons.

The opportunity for member organisations to exchange ideas and to discuss plans within the ESF Member Organisation Forum has been extremely helpful to sharing ideas in this field. We recommend that:

- Member organisations should continue to exchange views and good practice with respect to output collection;
- Research organisations support international efforts to uniquely identify researchers, and link this information to research output information;
- Output collection systems should work to align their definitions of data with the CERIF standard and provide CERIF compatible output;
- Rigorous methodologies for the valuation of output should be identified.

[1] The capture and analysis of research outputs
http://www.esf.org/index.php?eID=tx_nawsecuredl&u=0&file=fileadmin/be_user/CEO_Unit/MO_FORA/MOFORUM_Eval_PFR__II_/3rd_Workshop/Capture_and_analysis.pdf&t=1332584904&hash=147079e1e03c914e026e8229b966cec9ccf3e2b3

1. Background

1.1. ESF Member Organisation Forum on Evaluation of Publicly Funded Research

A number of organisations within the European Science Foundation Member Organisation (ESF MO) Forum on Evaluation of Publicly Funded Research[2] established a working group[3] to share their experiences and exchange expertise on the analysis of information about their research portfolios and also on the collection and analysis of output data. The working group has published two draft papers[4,5], largely drawing on the experience of organisations contributing to the working group.

In follow-up to the work on the collection and analysis of output data, the **MO Forum working group set out to provide further detail concerning the definition and analysis of key outputs**, and the ESF MO Forum undertook an international survey of research organisations to discover current practices. The aim was to share this information in order to help organisations develop their evaluation approaches.

ESF survey of international practice with respect to the collection of output data

The ESF survey was conducted in March/April 2011 and the working group held a preliminary discussion at the ESF MO Forum meeting in Paris (May 2011), a small follow-up meeting in London (October 2011), and agreed to prepare a new working paper at a second ESF MO Forum meeting in Bern in November 2011.

Thirty-two organisations responded to the survey and these are listed at Annex 1. All stated that they were actively collecting output data (27), or intended to (5). Those organisations intending to collect output data reported an interest in being able to evidence the value of the research they fund to politicians and the public, but also noted

[2] ESF MO Forum on Evaluation of Publicly Funded Research http://www.esf.org/activities/mofora/evaluation-of-publicly-funded-research.html

[3] ESF MO Forum on Evaluation Working Group current membership: Brendan Curran (Ireland, Health Research Board - HRB), Katharina Fuß (Germany, German Science Foundation - DFG), Iveta Hermanovská (Slovakia Academy of Sciences), Katrin Milzow (Switzerland, Schweizerischer Nationalfonds - SNF), Jenny Nordquist (Sweden, Swedish Research Council), Ingrid Roxrud (Norway, The Research Council of Norway), Ian Viney (UK, Medical Research Council - MRC), Rafael de Andrés Medina (Spain, Instituto de Salud Carlos III).

[4] The classification of research portfolios http://www.esf.org/index.php?eID=tx_nawsecuredl&u=0&file=fileadmin/be_user/CEO_Unit/MO_FORA/MOFORUM_Eval_PFR__II_/3rd_Workshop/Classification.pdf&t=1332584662&hash=792b3eddb558ce5caf485c454ebe07a4bdb9acd9

[5] The capture and analysis of research outputs http://www.esf.org/index.php?eID=tx_nawsecuredl&u=0&file=fileadmin/be_user/CEO_Unit/MO_FORA/MOFORUM_Eval_PFR__II_/3rd_Workshop/Capture_and_analysis.pdf&t=1332584904&hash=147079e1e03c914e026e8229b966cec9ccf3e2b3

interests in using the information as a tool for developing funding strategies, and for monitoring and assessing the development of individual projects and research fields.

The survey highlighted interest from a significant number of research funding organisations in the collection and analysis of information about research outputs, outcomes and impacts, to support their evaluation programmes. Research organisations are using this information to **better communicate the benefits of research to policy makers and the public, and to monitor and evaluate delivery against their strategic plans.**

The previous paper from this working group had highlighted:

- The need to ensure that, where possible, the administrative burden on researchers is minimised in order to maximise the time spent on frontline research.

- Some of the advantages and drawbacks to approaches for collecting output information. It was noted that organisations were planning and implementing different practices for the collection of output information, and so the report drew attention to consideration of the frequency of data collection, sources of data, and sanctions and incentives.

At the time it was highlighted that the following were areas of active discussion, and were only addressed briefly:

- The types of output/outcome/impact to be collected. **Although there is substantial agreement between organisations over established metrics such as publications and patents, there is a need for more effort to capture outputs that are more difficult to quantify, but give a more holistic picture of research output** (such as influences on policy and practice).

- Approaches for analysing output. **As research organisations develop more consistent and rich datasets on research output, better methods to track, visualise and assess the significance of these data are needed**.

This paper intends to **follow up on these points and provide more detail concerning the types of output data that are captured by research organisations and the ways in which this information is analysed and used to inform strategy development**.

1.2. Why collect and analyse research output?

Research funding and research performing organisations are interested in assessing and communicating the impact that research has on society, the economy or academia. Societal impacts might include the contribution that research, or researchers themselves, make to health and culture. Economic impacts include encouraging inward investment or improving productivity. Academic impact can include enabling further research by creating new techniques or training the next generation of researchers.

It is increasingly important to all funders to demonstrate that they can optimise the distribution of funding. Whether public funders accounting to Government, charities to their donors, research performing organisations to their sponsors, or companies to their shareholders, all organisations have to evidence that their strategies have added value, or that they can improve upon them.

More broadly, members of the public are interested in the results of research and have a right to know about the work they have paid for in the public and charitable sectors. Funding organisations are finding ways to provide open access to the results of the research that they support (e.g., DFG-GEPRIS - German Project Information System)[6] and improve the explanation of the ways that research leads to impact.

Approaches for measuring economic, societal and academic impact have been extensively discussed elsewhere[7,8]. The link between research and impact is often approached at a 'micro' level by studying hypothetical 'intermediate steps' in the pathway to eventual impact[9]. These intermediate steps might include the interaction between researchers and potential users or beneficiaries of research, for example via collaboration, commercialisation or more broadly via other dissemination routes (e.g., by publication or presentation at conferences, etc.). The hypothesis is that these activities may provide a leading indicator of research likely to produce impact. In order to measure these activities data are required about the volume, quality and diversity of these interactions. These data are usually referred to as 'outputs' and 'outcomes' from research, recognising that these activities are usually desired deliverables from research projects and programmes. The distinction between outputs, outcomes and impacts is

---

[6] GEPRIS: www.dfg.de/en/gepris

[7] *Measuring the link between research and economic impact* - report of an MRC consultation and workshop (MRC, 2012) http://www.mrc.ac.uk/Utilities/Documentrecord/index.htm?d=MRC008597

[8] *Measuring the Impacts of Federal Investments in Research: A Workshop Summary* (US National Academies of Science, 2011) http://www.nap.edu/catalog.php?record_id=13208

[9] For example see *Hidden Connections Knowledge exchange between the arts and humanities and the private, public and third sectors,* CBR report to the AHRC (2011) http://www.cbr.cam.ac.uk/pdf/AHRC_Report.pdf

often not clear, but what is important is that measures are found which provide good quantitative and qualitative indicators of progress. This might be formalised via approaches such as the 'payback framework'[10] which provides a multidimensional categorisation of benefits from research, starting with more traditional academic benefits of knowledge production and research capacity-building, and then extending to wider benefits to society.

This report is concerned with the approaches to gather output information, covering a range of different output types (section 2), general issues in gathering this data (section 3) and advice on potential approaches for the analysis of this information (section 4).

[10] The 'Payback Framework' explained, *Research Evaluation* (2011) 20 (3): 181-183. doi: 10.3152/095820211X13118583635756 http://rev.oxfordjournals.org/content/20/3/181.short

2. Types of output data

This section discusses several different types of output information, showing that each has different challenges for definition, validation and analysis. It is not suggested that all research organisations should analyse all these types of output, or should adopt particular definitions outlined. Each organisation may have different priorities for evaluation. However, we suggest that these outputs are important for policy evaluation and benchmarking and make a case for greater standardisation and consistency between organisations, so that wider joint analyses across organisations may be possible. Examples of output and outcome information are included to illustrate some current practices and promote discussion between research organisations about evaluation approaches best suited to improving strategy development and communication of impact.

Some of the output types chosen are collected by organisations across Europe (e.g., publications, intellectual property) and could in principle be analysed across these organisations (for example, see results from the ESF MO Forum on Evaluation: Indicators of Internationalisation[11])

2.1. Publications

Publications are by far the most commonly analysed output from research, largely because it is one of the most direct outputs, but also because international data are available through publishers. Publishers' databases enable large-scale bibliometric analyses, which may serve as an indicator for the production (number of articles) and scientific use of research results (citations). However, these databases are mainly focused on peer-reviewed articles and therefore have a rather weak coverage in, for example, the areas of humanities and social sciences where monographs are the more common way of communicating research results. They are also limited in classifications applied, which are mainly disciplinary, and in correctly attributing articles to specific research funding or research performing organisations.

As a result of these limitations, it is becoming increasingly common for funders to collect publication data through electronic reporting systems at the end of a granting period. This approach means completely new possibilities for detailed output analyses, including a variety of publication types and classifications as well as improving the accuracy of linkage to funding inputs. However, it also means that new local databases are created,

[11] This MO Forum is undertaking a pilot study aiming to design and to produce a set of indicators that could account for the internationalisation of European research activities and programmes. http://www.esf.org/activities/mo-fora/evaluation-indicators-of-internationalisation.html

and these may diverge and miss the opportunity for wider analysis across Europe and internationally, if they do not follow international standards and common data structures.

2.1.1. Researcher identification

One major problem with publication data today is to accurately connect author names to one individual, and to connect individuals to institutions. This is due to misspellings, change of names, use of different acronyms for organisations, etc. There are initiatives to this end (e.g., researcher ID[12] and ORCID[13]) but still no unified standard has been accepted. The working group agreed that it was important for research organisations to support international efforts to uniquely identify researchers, and link this information to research output information.

2.1.2. Publication types

While there is a lot of focus on journal articles across different disciplines, the importance of this publication type varies. In the humanities and social sciences, for instance, monograph publications are among the most important ways of communicating research results. Research organisations therefore need to ensure that output collection systems are designed to be able to accommodate a wide range of publication types. This might include: articles (original, review, proceedings) – published or accepted for publication; books, contributions to books, edited volumes; scientific presentations (oral presentations, posters); wider audience (seminars, lectures, exhibitions, public media).

2.1.3. Publication data fields

The data fields required by funders include: type and status of publication, for articles (name of author(s), title, journal name, year, volume, issue, page numbers, abstract), for books (name of author(s), year, publisher, ISBN-number), digital object identifier (DOI), open access status.

Many funders help the user by importing data from publishers or publication repositories to pre-populate their output collection system. Users may be presented with papers that have a match to their researcher ID, grant or surname and initials. Users are then asked to validate this information by claiming papers. Output collection systems usually allow input of an identifier which is then used to import bibliographic data from publishers or repositories. Data imported from publishers will include: name of author(s), authors' addresses, year of publication, title, source (publisher), abstract, keywords, reference list (citations) and acknowledgements. The data may also include classifications: either

---

[12] Researcher ID is an online initiative to uniquely identify researchers. Researcher ID is integrated with Thompson Reuters Web of Science allowing authors on papers to be specifically searched for and citation metrics to be compiled. http://www.researcherid.com

[13] ORCID (Open Researcher and Contributor ID) aims to solve the name ambiguity problem in scholarly communications by creating a registry of persistent unique identifiers for individual researchers and an open and transparent linking mechanism between ORCID, other ID schemes, and research objects such as publications, grants and patents. http://about.orcid.org/

publicly available categorisations (such as MeSH[14] applied by PubMed[15]) or proprietary data (such as Thompson Reuters journal classifications).

2.1.4. Analyses

As mentioned above, it is important that research quality is analysed and evaluated in a way which best reflects quality in a particular field of science. Analyses of research results should therefore include publications of relevance, not necessarily be limited to journal articles. Comprehensive databases of citations outside of the life sciences have more recently become available[16], although bibliometrics is largely utilised by the STEM subjects.

Usually information about publications is the most significant part of the output from research projects. For example, despite examining many other output types both the Medical Research Council (MRC) and the Swiss National Science Foundation (SNF) have found that around half of the reports collected from their funded research groups are journal articles or other types of publication.

Bibliometric analysis provides the most extensively benchmarked data about quality, and this is often relied upon to assess the relative impact of specific research programmes. However, the significance of citation impact data is poorly understood. Very highly cited publications are rare, and so if small numbers of papers are analysed the results are likely to be 'noisy'. It is helpful to look at, and contrast, distributions of citation impact (c.f. Thompson Reuters 'impact profiles'[17]) in order to assess whether a particular population of papers has overall higher citation impact.

The lag time between research funding and resultant papers achieving citation impact should also be considered. The issue of lag time is important, if calculations of productivity are to be attempted. Lag time will of course vary by discipline, scientific field, type of research activity, etc.

---

[14] http://www.nlm.nih.gov/mesh/

[15] http://www.ncbi.nlm.nih.gov/pubmed

[16] Thompson Reuters have released an Arts and Humanities citation index http://thomsonreuters.com/products_services/science/science_products/scientific_research/scholarly_information_journals/arts_humanities_social_sciences/arts_humanities_citation_index/ , and a Social Sciences citation index http://thomsonreuters.com/products_services/science/science_products/a-z/social_sciences_citation_index/ . The ESF has worked with Elsevier to extend its coverage of the Arts and Humanities http://www.esf.org/research-areas/humanities/news/ext-news-singleview/article/esf-works-with-scopus-to-expand-arts-and-humanities-coverage-581.html and this is also an issue that the DFG have examined http://www.dfg.de/download/pdf/foerderung/grundlagen_dfg_foerderung/informationen_fachwissenschaften/geisteswissenschaften/annex_2_en.pdf

[17] Profiling citation impact: A new methodology, *Scientometrics* Vol. 72, Number 2 (2007), 325-344, DOI: 10.1007/s11192-007-1696-x

The full-text content of papers can also be analysed to follow the evolution of ideas. This approach was used by Pierre Azoulay and others[18] in 2010 to examine the extent to which researchers explored ideas outside of their original field of interest. The work took the novel approach to try to analyse quantitatively the extent to which funding schemes gave researchers 'freedom to explore' (comparing HHMI researchers to NIH grant holders).

All this detail about the quality of output, location and identity of collaborators and development of ideas is used by many research organisations to monitor, review and communicate progress.

2.1.5. Trends for capturing and analysing publication output

As publications are the most common type of output most output collection approaches have tried to make the entry of these details as easy as possible. The existence, at least for biomedicine, of comprehensive publication databases means that records of papers with accurate bibliographic details can be linked to.

The accelerating move to open access publication[19] means that more extensive information about publications will in future be freely available. Full text searching of papers will be helpful in analysing the use of different methods and datasets in research, tracking the evolution of ideas, and extracting acknowledgements.

Over the last five years UKPMC[20] has grown from a simple mirror of the USA-based National Center for Biotechnology Information PubMed Central site to a stand-alone site providing access to a repository of more than two million full-text biomedical research articles, more than 25 million citations from PubMed and Agricola, patents from the European Patent Office, UK treatment guidelines and biomedical PhD theses. The UKPMC funders group has been joined by the Austrian Science Fund (FWF) and Telethon (Italy), and in July 2012 it was joined by the European Research Council[21]. As a result of this participation, the existing funders have agreed that the service will be rebranded as 'Europe PubMed Central' (Europe PMC) by November 2012. A key aim of this initiative is to extend the repository further and encourage other European funders of life sciences research to make the outputs of the research they fund freely available through Europe

[18] Incentives and Creativity: Evidence from the Academic Life Sciences, *NBER Working Papers* 15466 (2010) http://pazoulay.scripts.mit.edu/docs/hhmi.pdf

[19] "Open access is the future of academic publishing, says Finch report" (UK Guardian Newspaper, June 2012) http://www.guardian.co.uk/science/2012/jun/19/open-access-academic-publishing-finch-report

[20] UK PubMed Central is a free digital archive of biomedical and life sciences journal literature, set up by a group of UK research funders, led by the European Bioinformatics Institute in partnership with the British Library and the University of Manchester. http://ukpmc.ac.uk/

[21] European Research Council renews its commitment to open access by joining Europe PubMed Central http://ukpmc.blogspot.co.uk/2012/07/european-research-council-renews-its.html

PMC. It is clear that a truly European PubMed Central resource would be of benefit to European research evaluation.

As with other outputs, standards for the mapping and exchange of publication data would help link sources of data on papers with information about funding, impacts and researchers.

2.2. Intellectual property and routes to commercialisation

Research funding agencies are increasingly interested in tracking intellectual property[22] developments and pre-commercial outputs of the research they fund, particularly where they have introduced strategic funding programmes targeted at the development of such outputs (e.g., translational research programmes). Moreover, there is increasing pressure on funding agencies from government budget-holders to demonstrate economic and commercial outputs from publicly-funded research and to show how this research is contributing to economic, cultural and social development. Given the typical short-term expectations of government and policy makers in terms of economic impact from research, it is important that funding agencies capture early outputs of the research they fund in relation to intellectual property rights (e.g., patents) as indicators of potential commercial and economic impact. Furthermore, they should continue to track the commercial exploitation of this IP through spin-outs, licence agreements and so on as it hopefully manifests into real commercial and economic impact.

IP	Utility patent (technical)	Design patent	Trademark	Copyright	Trade secret	Registered design
Protection	Products, devices, processes, novel application	Industrial design	Words, logos, symbols	Expressions of original creative works, training manuals	Confidential information kept secret	Industrial designs
Term	20 years	14 years	As long as it is used	70 years minimum	As long as secret is maintained	10 years
Registration	Yes	Yes	Not required (but possible) ™ or ®	No (but possible) ©	No	Yes
Costs (to obtain and maintain)	High (up to 20k over lifetime)	Medium	Low	Low	Low	Low

Table 1 Types of intellectual property protection

Table 1 outlines the main types of intellectual property protection, contrasting the kinds of discovery protected and process required to put this protection in place.

2.2.1 Findings from MO Survey

What sort of IP data is collected by MOs and how is it collected?

The main Forum survey of the member organisations showed that 23 organisations (85%) collect IP data, mainly via annual/end-of-grant reports (most common currently)

[22] See glossary for an explanation of terms such as 'intellectual property'.

or online output collection systems (increasingly common). Most agencies ask quite general questions in this regard (e.g., "*Has any intellectual property arisen as a result of research carried out in this project*?") and request some basic information such as type of IP, title (of patent), details of inventors and owners, etc. Some agencies request more detailed information (such as whether the IP has been licensed, and if any subsequent impacts of the discovery are known) (e.g., MRC – see below), and some are planning to do so, particularly as they develop more sophisticated online outputs reporting systems (e.g., DFG).

Information is sought on the following areas (aggregated across agencies):

- Engagement with university technology transfer office (TTO)
- Invention disclosures submitted to TTO
- Patents filed/applied for – details (e.g., national/EU/US/other)
- Patents awarded/registered/published – details (priority date, inventor(s), patent title, abstract, jurisdiction, patent number)
- Patents lapsed (i.e., maintenance protection fees not paid)
- Any copyrighted material
- ICT standards/specifications
- Industrial designs and trademarks
- Confidentiality agreements to protect 'know-how'
- Cases where IP protection is not required.

Information also commonly sought on commercialisation routes:

- Licence agreements with third parties – details (e.g., commercial / non-commercial; value of royalties)
- Industry partnerships (through collaborative research and shared IP)
- Spin-out companies formed – details (e.g., date of incorporation, further funding from venture capital or technology development grants, jobs created).

What analyses are carried out on IP data by MOs?

Generally speaking, the analyses carried out by agencies are very basic descriptive statistics such as count of patents, aggregated value of licence agreements, number of spin-outs, etc. In some cases, IP outputs (as per other outputs) are aggregated to the level of funding programme / scheme for evaluative purposes.

Some agencies such as the MRC carry out more in-depth analyses, for instance by differentiating between licensed versus unlicensed patents (as an indicator of market potential) and comparing to international data.

In 2010 the MRC showed that approximately 30% of patents arising from MRC-funded research since 2006 were reported as licensed by 2010, while a study of over 1,200 patents published by the Universities of California and Columbia found that 41% of patents were licensed by 1999[23]. A similar study of 686 patents published by the Memorial Sloan-Kettering Cancer Centre and Dana Faber Cancer Institute between 1983 and 2003 also found that 41% of these were licensed by 2007[24]. Other studies have indicated a lower proportion of patents licensed (for example, 25% of NASA patents published between 1994 and 2002 were licensed by 2007[25]).

What are the analyses used for?

Data relating to IP and technology transfer have many potential uses for agencies, in terms of both internal (e.g., strategic) and external (e.g., accountability, advocacy) audiences. Currently, MOs reported the use of IP data at two broad levels:

Individual grant level

- Source for assessing the progress/success of individual projects – particularly for those projects in the translational space or research expected to produce IP (e.g., translational research awards);
- To feed into general evidence pool to assess past performance, if the applicant reapplies for funds;
- Making the results of funded research available to the public.

Aggregated level

- Assessment of the relative quality of research funded (e.g., across an entire portfolio);
- Evidence source for evaluation of funding schemes and programmes, for strategic decisions;
- Justification of funding to political circles and the public (corporate annual/performance reports, outcome compendiums for the public, website);
- Evidence of an agency's impact for government spending reviews (e.g., as used by MRC in UK and HRB in Ireland).

[23] Patent citations and the economic value of patents – a preliminary assessment, Sampat and Ziedonis, Chapter 12 in *Handbook of quantitative science and technology research: the use of publication and patent statistics in studies of S & T systems* (2004) http://www.springerlink.com/content/j40012303765345m/

[24] *Patent citations and licensing value* by Gregory P. Daines, MBA dissertation (Massachusetts Institute of Technology, 2007) http://dspace.mit.edu/handle/1721.1/39530

[25] *Federal Lab Patents, Licensing, and the Value of Patents: Exploring the Licensed Patents from NASA* (T. Jung, Georgia State University 2007) http://www.spp.gatech.edu/faculty/WOPRpapers/Jung_WOPR.pdf

2.2.2 EU Guidelines on Knowledge Transfer Metrics

Agencies may wish to consider the report[26] of an Expert Group on Knowledge Transfer Metrics set up by DG Research of the European Commission to improve coherence and convergence between existing surveys of knowledge transfer from public research organisations (PROs) to business and other sectors in society. The purpose of this report was to improve the possibility for individual PROs and Member States to monitor and compare their achievements in this field against themselves over time and against each other, in order to identify trends and to support work on improvements if needed.

Recommended core performance indicators for the PROs:

- Research agreements
- Invention disclosures
- Patent applications
- Patent grants
- Licences executed
- License income earned
- Spin-offs established.

Suggested supplementary indicators for more detailed monitoring of the core performance indicators:

- Knowledge transfer involving small- and medium-sized enterprises (SMEs)
- Knowledge transfer involving domestic firms
- Knowledge transfer involving the research organisation's own region
- Exclusive licences
- Share of valid patent portfolio that has ever been licensed
- Patent share of licence income
- Technology areas for patenting.

2.2.3 Are there other sources of IP data that funding agencies can utilise?

Agencies may also wish to consider other repositories of IP data that may contain potentially richer (and more accurate) information in terms of either collecting the data at source or verifying IP data collected via grant reports or PI outputs surveys.

Some of these potential sources are listed below.

- University Technology Transfer Offices – typically have systems to record details of all commercialisation activities from invention disclosure through to licence

[26] *Metrics for Knowledge Transfer from Public Research Organisations in Europe: Report from the European Commission's Expert Group on Knowledge Transfer Metrics*. European Commission (2009)

agreements and/or spin-outs established. Records would include funding inputs to any IP.

- National Patent Offices - maintain a database of all national patents filed and awarded and typically publish details on their website and in annual reports.
- European Patent Office (www.epo.org) – provides free access to the European Patent Register (all EU patents) and Espacenet (a repository of 70 million patent documents worldwide containing information about inventions and technical developments dating from 1836).
- Commercial repositories, such as the Global Patent Index (see EPO website above) and Thomson Reuters Derwent World Patent Index.

The MRC carries out some verification of its patent data collected through e-Val by matching it against the UK Intellectual Property Office (IPO) dataset. Thus, in 2010, 225 reports of patents were matched to the IPO dataset, duplicates removed and publication dates validated/amended. This resulted in further validated information on 185 reports of patents published since 2006[27].

2.3. Products and interventions

When an application of research is identified with potential positive impact, efforts should be made to introduce this into practice. This usually involves protecting the idea (see section 2.2 above), and then seeking to find a route to exploit it (for example by commercialising it, or working with others to disseminate and drive adoption). This may entail further development of the product or intervention before it is a commercially attractive proposition, or before there is enough validation/further evidence for it to be adopted into practice.

Products and interventions will progress along a developmental pathway toward adoption, and if research organisations are interested in the time taken to translate ideas into practice then easily defined stages are needed along this pathway.

The US Army Medical Department has outlined how pharmaceutical products map to 'technology readiness levels' which are used to assess the maturity of a technology[28] and are widely used in engineering and other disciplines. This provides a ten step pathway from fundamental research to wide scale adoption.

---

[27] http://www.mrc.ac.uk/Achievementsimpact/Outputsoutcomes/MRCe-Val2010/Intellectualproperty/index.htm

[28] Table 2 – Technology Readiness Level Descriptions for A Pharmaceutical Product https://mrmc.amedd.army.mil/index.cfm?pageid=researcher_resources.ppae.atostat

Comprehensive information on the types of product or intervention in development, and the stage that work has reached, should be a powerful tool for research organisations in understanding the points of handover (when product development passes to commercial partners, or collaborators), and helpfully identifying barriers to translation.

Very few organisations currently capture this information[29].

[29] MRC e-Val data on products and interventions http://www.mrc.ac.uk/Achievementsimpact/Outputsoutcomes/MRCe-Val2010/Productsandinterventions/index.htm

2.4. Influence on policy

Research as a tool to meet major future challenges of our societies is gaining increased attention, and the public sector plays a key role in meeting these challenges. The public sector is in need of research-based knowledge, e.g., for developing and improving policy and services. Innovation in the public sector is seen as an important tool to renew and develop the sector. The dynamics of the public sector differ from the private sector in several respects, and therefore differentiated strategies for research-driven innovation may be required.

Being able to capture data related to influence of funded research on the public sector can therefore be very useful for research funders and performers, generally or in the context of specific funding instruments or scientific disciplines/fields (e.g., health care services research). A majority of the organisations responding to the survey report using output data for strategy development and to evidence funded research value towards political actors and public opinion, and this output type can be valuable to this end. Furthermore, in evaluations of, e.g., funding schemes where the public sector is envisioned as an important user of the research results, or studies aimed at understanding how research impacts our societies, these data might prove useful.

Several output types can be appropriate in order to gather information on how research influences the public sector. Influence on policy is one of these, and 33% (9/27) of the survey respondents report collecting this type of output data. Other output types elaborated on in the report can also be important with regards to influence on the public sector. Examples are Products/Interventions (see 2.3) collected by 59% of the survey respondents, Other dissemination of knowledge (see 2.7) collected by 74% and Collaborations (see 2.6) collected by 74% of the respondents. For these output types one might consider differentiating between outputs relevant for/aimed at the private sector, the public sector, the academic world or other. Including additional output types might also be useful, for example participation in committees appointed for policy-related tasks.

To measure outputs with influence on the public sector is difficult, probably more so than commonly used output types such as publications. The time lag between the research performed and the effect on the public sector is one very important issue. Information about the influence that researchers have had on the setting of new policies is often (though not always) captured years after awards have finished.

A recent study examined the funding sources for research that had been cited in UK clinical guidelines[30]. This study showed that it was feasible to identify in a semi-automated way the research that had been used to evidence influential policy documents. The results also showed that the lag time between publication of research papers and citation in the clinical guideline, in the two specific cases analysed, was three to five years. This approach is only feasible where policy documents take care to cite all the evidence used in drawing up recommendations, and it may still be resource-intensive to extract all the relevant data. There are over 200 guideline issuing organisations in clinical medicine noted by the US clearing houses for guidelines[31], and many other standards and policies relevant to other disciplines worldwide.

Researchers may not know if their research has been cited in a policy document, and so be unable to report this to their research organisation. It is more likely that they will be able to report substantive influences, particularly where their research can be shown to have had an impact on specific recommendations. This information may be more immediately useful to research funders than capturing all citations in policy documents.

Due to the time lag, and the fact that a number of research awards are likely to have contributed to the evidence for policy recommendations, attribution will be challenging. This is the case with many other output types.

In a context where gathering information about the influence of research on the public sector is important, the involvement of stakeholders in the public sector, the 'users' of the outputs in question, may be interesting as a part of assessing the reach and significance of the reported output data.

[30] Tracking the impact of research on policy and practice: investigating the feasibility of using citations in clinical guidelines for research evaluation, *BMJ Open* 2012;2:e000897 doi:10.1136/bmjopen-2012-000897

[31] http://guideline.gov/

2.5. Mobility, Next destination, Training

Researchers' geographic or institutional mobility, training opportunities and subsequent career paths are areas of common interest for many funding agencies. Research organisations invest significant funding to support young researchers at PhD or postdoc level, and the supply of highly qualified personnel to the world economy is considered an important aspect of the impact from research. Researchers' mobility and career progression are, however, challenging to capture, analyse and benchmark. Most approaches rely on being able to maintain contact with people who have benefitted from research funding, despite the fact that they may be highly mobile (career moves may take them anywhere internationally), and they may pursue careers outside of academia and even outside of research. In addition, relevant changes may only occur over a long period of time. Lastly, analysis of this data is hampered by the lack of widely accepted definitions or taxonomy for elements such as 'career stage'.

2.5.1. Output data surveys

Nine of the 32 organisations surveyed by the working group, all of which are research funding organisations, collect or intend to collect data on the next destination of staff, especially at a junior level (PhD, postdoc). Eight institutions collect data on skill shortages/training issues (six funding organisations, two research organisations). Next destination and training belong to the less frequently collected output types, but are still gathered by a sizable minority, which provides some scope for a comparison of approaches.

Data collected with output data surveys or end-of-project reports can provide information on the number and categories of staff funded, as well as the role, sector or geographic location taken on by researchers after their funding. The questions of analysis addressed in this context converge on the proportion of staff in employment, staff active in the academic sector, or in a sector of particular interest (for example, the health sector), which are drawn on in corporate reports or communication with political players and the public to assess and demonstrate impact. Organisations also draw on the data for analytical and strategic purposes, for example to identify shifts in research fields or, combining destination countries with research areas, emerging countries in different fields. Thus, information about the location of the next destination is used to define strategies promoting internationalisation and mobility.

> "62% of staff who left (…) remained in the academic sector (…) 11% moved into the private sector" (MRC)
>
> "99% of 106 postgraduate students were in current employment.10% were based outside Ireland" (HRB)

Although there is significant interest in similar questions for analysis, data collection practices (end-of-project report, ongoing data collection) as well as definitions of categories (for example, sectors or roles) vary considerably, so it is not evident as to how results can be benchmarked. The lack of a widely accepted taxonomy for career stage has been recognised by the ESF MO Forum on Research Careers. This forum has recommended in its final report[32] that a working group be established between ESF member organisations, the European Commission and universities in Europe (represented by the European University Association and the League of European Research Universities) to work out a joint taxonomy to better orientate evaluation of research careers.

A more fundamental shortcoming of data collected within the scope of output data surveys is that it shows a snapshot at a particular time and does not provide a long-term perspective on researchers' career development. Since the unit of analysis in output data surveys frequently is the project, and the next destination of project staff is, unlike for example publications, not cumulative, even repeated or opened ended surveys allowing researchers to compile outputs as they arise are not much help in this regard. At most, open ended surveys can reduce the dependence of results on the timing of data collection, if the next destination is not immediately known.

2.5.2. Career tracking

There are several longitudinal studies underway that aim to look at careers over the longer term to understand researchers' career pathways. These studies can take the form of prospective cohort studies, panel studies or surveys tracing back careers over several years. The field was examined in a joint ESF-FNR workshop in February 2012[33].

The questions of analysis addressed by such studies focus on issues similar to those addressed in output data surveys. Employment situations and destinations (percentage employed, in academia, in similar research field) are a key concern in the OECD/Eurostat Careers of Doctorate Holders Project[34], the EMBO report *A persistent problem*[35], the Swedish Research Council (VR) report *Career development and success*[36], and in the Wellcome Trust Career Tracker[37], a longitudinal prospective study to track Trust award

[32] *Research Careers in Europe Landscape and Horizon* (ESF, 2012) http://www.esf.org/fileadmin/links/CEO/ResearchCareers_60p%20A4_13Jan.pdf
[33] International workshop 'How to Track Researchers' Careers' (ESF/FNR, 2012) http://www.researcherscareers.eu/index.php?p=home
[34] Careers of Doctorate Holders Project http://www.oecd.org/sti/cdh
[35] *A persistent problem*. Anna Ledin, Lutz Bornmann, Frank Gannon and Gerlind Wallon, EMBO reports Vol. 8, no. 11, (2007)
[36] http://www.vr.se/download/18.7f636125fc6308e48000449/1262869674523/Career+development+and+success_5+2009_FINAL.pdf
[37] Wellcome Trust Career Tracker http://www.wellcome.ac.uk/Funding/Biomedical-science/Research-careers/WTDV026334.htm

holders' careers over time. Analysing mobility in and out of higher education employment over a three-year period, a longitudinal destinations survey of more than 2,000 doctoral graduates by Vitae[38] illustrates the added value provided by longitudinal studies,[39] as do results from the Wellcome Trust Career Tracker showing that the proportion of men staying in academia is fairly consistent, while the proportion of women decreases.[40]

> "Over 3 years, 40% of the cluster moved out of higher education research occupations (11% to HE teaching); 26% moved into higher education research occupations" (*What do researchers do?* Vitae, 2011)
>
> "30% of researchers that published work while affiliated with UK institutions stayed in the UK for less than two years before moving abroad" (DBIS, Elsevier, 2011)

Longitudinal studies also provide scope for further-reaching analyses about the impact of training programmes than data collected within typical output data surveys. For example, project based output data can show how long funded PhD candidates took to complete their PhD; tracking their careers may provide insight into how this impacted their research career in the long term. Providing a foundation for improving the design and implementation of PhD programmes is a key concern for the Profile project conducted by the German IFQ[41] in collaboration with the DFG, which surveys PhD candidates at the beginning and end of PhD, as well as four years after PhD completion. The employment situation after the PhD is then analysed in consideration of individual characteristics and the structure of the PhD programme to gain insights into the repercussion the structure of the programme has on the development of PhD candidates.

In addition to career tracking through surveys, there are some initiatives relying on existing data. STARMETRICS[42], a project led by the NIH, NSF and the White House Office of Science and Technology Policy in the US, aims to address similar questions relying directly on data from research institutions' existing database records, which are to be combined with existing publication, citation or patent data. With respect to mobility and career paths, STARMETRICS aims to analyse job creation as well as workforce outcomes like student mobility and employment. Also relying on existing data, a report commissioned by the UK Department of Business, Innovation and Skills on the

---

[38] *What do researchers do? Career paths of doctoral graduates* (Vitae, 2011) http://www.vitae.ac.uk/CMS/files/upload/Vitae_What_do_researchers_do_Career_paths_2011.pdf

[39] *What do researchers do?*, Vitae, 2011, p. 4

[40] Wellcome Trust Career Tracker, Presentation at the joint ESF-FNR Workshop 'How to Track Researchers' Careers', 9-10 February 2012, Luxembourg.

[41] IFQ Profile project http://www.forschungsinfo.de/profile/start.html

[42] STARMETRICS (Science and Technology for America's Reinvestment: Measuring the Effect of Research on Innovation, Competitiveness and Science), is a multi-agency venture led by the National Institutes of Health, the National Science Foundation (NSF) and the White House Office of Science and Technology Policy (OSTP). https://www.starmetrics.nih.gov/

International Comparative Performance of the UK Research Base - 2011 adopts an altogether different approach to analysing mobility: on the basis of publication data and authors' affiliations listed in their articles, this study quantifies the flows of researchers, or brain circulation, in and out of the UK over the period 1996-2010. It finds, for instance, a net brain outflow from the UK of about 1.5%, where the inflow groups, however, tend to be more productive than the outflow groups.[43]

Several research organisations have used the curriculum vitae (CV) as a useful format for gathering and re-using output data. As researchers regularly use CVs to apply for posts, and set out their track record and career history in grant applications, this may be an efficient approach. This is the basis of the successful Lattes platform[44] implemented in Brazil. Lattes is a national database of CVs which funders require researchers to update. These CVs are then used in grant applications, recruitment, etc. The Lattes platform allows researchers to easily drag information from publication databases into the online CV. The US STARMETRICS programme has expressed an interest in Lattes as a potential route to populating Level II data (information about outputs and outcomes), although this would require the regular updating of CVs and the careful expansion of these documents to include information about the progress of projects.

Approaches relying on existing data still raise the question of how different data sources are to be combined. Of particular concern to research funders is the association between researchers, research groups/organisations and funded projects. For some issues of direct concern, such as the immediate next destination after funding, output data surveys may provide a good alternative. When designing questions addressed with output data, organisations should, however, be careful to limit themselves to issues that can be addressed within the 'snapshot' framework of project reporting and exploit synergies with longitudinal studies. On the other hand, existing output data may provide an input when designing future career tracking initiatives. Ideally, 'output data systems' linked to both projects and persons could provide a holistic picture of how several research careers intersect to contribute to a research line.

---

[43] International Comparative Performance of the UK Research Base: 2011, Department of Business, Innovation of Skills, Elsevier, 2011
[44] http://lattes.cnpq.br/english/index.htm

2.6. Collaboration

Collaboration as a scientific output has not often been discussed or analysed for several reasons. First of all, there is no common accepted definition of collaboration as one type of scientific output. Furthermore, there are no common indicators to measure the effect of collaboration. Thus, the understanding of collaboration can be different in every funding or research organisation. However, it is worth analysing this output type as knowledge grows by sharing - by working together with people from other institutes and fields of research, on a national as well as on an international basis.

Collaboration usually includes teams working together to achieve common goals, sharing facilities, resources and expertise. New knowledge evolves by cooperating and communicating in various ways. In particular, the collaboration between researchers of different disciplinary origin or nationality can lead to innovative methods and results. Funding organisations encourage collaborations not only because of the knowledge transfer among researchers or research organisations. The knowledge transfer between science and the economy also plays an important role.

Research organisations are interested in the added value that collaboration brings. Many studies have shown that publications with higher citation impact arise from collaborative work. Some of this impact may be due to a larger network of researchers participating in onward communication of the work, but the impact appears to increase with cross-sector collaboration (e.g., industry/academia interaction) and international collaboration, not simply numbers of authors.

This has led to research funders launching dedicated schemes to encourage collaboration[45]. Thus funders are interested not only in demonstrating the benefits of supporting such interactions, which evaluations have begun to demonstrate[46], but also in the conditions under which incentives to encourage cooperation are likely to be effective. If, from the first angle, cooperation may be considered an input more than an output, from the second angle, it is very much an output of funding activities.

In the survey of ESF member organisations, three out of four participants indicated that they collected data about collaborations. The ways to collect these data are various, so are the approaches to operationalise this output type.

---

[45] The DFG, for example, funds Transfer Projects in Collaborative Research Centres. This programme element addresses projects designed to test the results of basic research under practice conditions or to translate them into prototypical applications. These projects generally involve collaboration with external partners.

[46] An example being the PASEC evaluation of the UK technology Strategy Board Collaborative Research and Development Programmes (2011) http://www.innovateuk.org/_assets/pdf/publications/pacec_evaluation_of_crandd_report.pdf

Some organisations report the number of collaborations they maintain. Before interpreting such a number and comparing it to the figures of other organisations, it is important to have a closer look at what falls within the concept of collaboration. Some additional attributes about the collaboration partners and the fields of collaboration should be recorded to prevent comparing apples with oranges. It should be possible to filter the data by the type of collaboration partners (e.g., academic, non-academic, institutional, with single researchers, national or international) and the fields of collaboration (e.g., exchange of personnel, joint publications, training for doctoral students, joint conferences, in-depth/constructive exchanges on approaches, methods, infrastructures or results).

End-of-grant reports and monitoring reports during and beyond the funding period are often used as sources for data about collaborations but it is also possible to use bibliometric databases to analyse co-authorships. Publications can provide good evidence of collaboration. While not all collaborations may result in published results, co-authorship is often used as a measure of these interactions. By extracting information on the addresses of authors, analysis of co-authorship across locations and sectors (e.g., public, charitable and private sectors) can be performed. If authors can be specifically identified then analysis at the individual level can be carried out.

Excluding bibliometrics, the analyses of the data are mostly descriptive statistics, as for example the breakdown of the number of international collaborations according to the resident country of the partner. This empirically based information is primarily used to disclose what research funds provided are used for. The analyses are mostly used to comply with the request for transparency concerning the project results as the demand for funding organisations to justify their expenditures to political audiences and the public grows ever stronger.

Looking across the practices of six funding organisations - German Research Foundation (DFG), Health Research Board of Ireland (HRB), UK Medical Research Council (MRC), Research Council of Norway, Swiss National Science Foundation (SNF) and the Swedish Research Council (VR) - Table 2 shows what sort of data on collaborations funders collect and how. The forms of analysis and contexts in which the organisations use the results are also presented.

As many organisations have just, or are in the process of developing new IT systems for collecting output data, it is the right time to establish a structured and standardised way to ask questions about collaborations – thereby getting more valuable and useful data for understanding how cooperation comes about, how research funders can support them and how they impact on research.

Table 2 A comparison of the data selected organisations gather about research collaborations

	What sort of data does the organisation collect?	How does the organisation collect the data?	What analyses does the organisation carry out with the data?	What are the analyses used for?
DFG	Type of collaboration partners: - academic and non-academic institutions (national and international by country) Field of collaboration: - training for doctoral students - joint conferences - exchange of academic staff - joint publications	- Final Report (so far PDF but currently developing a web-based, structured solution) - annual monitoring survey (Collaborative Research Centres and Research Training Groups) - one-off evaluation studies	- descriptive statistics - network analysis - bibliometric analysis/ co-authorship (one-off studies)	- accountability/ transparency about what the funds provided are used for - political advice - basis for planning purposes - evaluation of funding programmes On a project basis: - source for assessing the success of individual projects
HRB	Type of collaboration partners: - industry-academic - academic to academic - international industry collaborations (by country) Field of collaboration: - joint publications - joint events (international) - networks	- End-of-grant report	- descriptive statistics - bibliometric analysis/ co-authorship	- the data are used for the report *Analysis and Outputs of HRB Grants* - baseline for future output reports for monitoring trends in research output
MRC	Type of collaboration partners: - collaborations between single researchers - collaborations with partner organisations - represented countries among these partnerships	- MRC e-Val: Researchers are asked to report details of each partner involved in their collaborations (annual survey)	- descriptive statistics - most frequent partners reported by location country - partners by sectors - interaction with the private sector - average number of partners in a single collaboration	- collaboration and partnerships play an increasingly important part in MRC`s mission to improve the health and wealth of the nation

Research Council of Norway	Type of collaboration partners: - professional or financial resources	- project applications - information on partners should be updated throughout the project period through the annual progress reports	- descriptive statistics	- analyses and use of these data as an output is mainly carried out when such information is requested, e.g., by universities - grant proposal assessment process - intention to be able to link collaboration to funding in the future
SNF	Type of collaboration partners: - academic and non-academic institutions - country of partner institution Fields of collaboration: - in-depth/constructive exchanges on approaches, Research Infrastructures, methods or results - joint publications - exchange of personnel	- online grant administration system	- descriptive statistics	On a project basis: - source for assessing the success of individual projects - making the dynamics of funded research visible to the public Aggregated data: - mapping of research processes and dynamics - assessment of the relative quality of research funded - assessment/reconfiguration of funding instruments - strategic decisions - justification of funding to political circles and the public
VR	- information about co-applicants - co-authorship	- VR is developing a new IT system to collect output data - Bibliometric analyses: VR buys data from Thompson Reuters yearly and stores it in a database set up in-house	- bibliometric analyses - level of collaboration (in different fields and over time) - collaboration networks - multidisciplinary - subject profiles for different countries or universities	

2.7. Dissemination of results to non-scientific audiences

Research organisations are interested in the ways in which researchers disseminate the results of their work. A primary way in which results are disseminated is of course via the publication of articles, but there are many other ways in which the outputs from research can be communicated.

A significant part of the research process is the discussion of ideas and progress with scientific peers, and this activity is pursued informally (every day within research organisations) and formally (via scientific meetings) as part of the work of every researcher. Capturing the myriad of interactions within the scientific community may not be feasible given the high volume of these. While literature, such as conference proceedings, and electronic links such as social media could be analysed, generally the extent of scientific collaboration is usually investigated using evidence from co-publications or surveys designed specifically for this purpose.

Given that research is a highly interactive process, attention is often focused on the way that researchers interact with 'audiences' for their work outside of the scientific community, e.g., in the media or lectures open to the public. To our knowledge there are no systems to record these interactions, and so output collection approaches are needed to capture this information.

> The MRC e-Val[47] approach is used by the UK Medical Research Council to record dissemination activities that researchers have engaged in. Researchers are asked first to report the format that the engagement activity took ("a visit to my laboratory", "a talk/presentation", "a formal workshop", etc.), and are then asked for the primary audience for the activity ("school children", "policy makers", "the general public", etc.). Details are then gathered about the activity including any data about the impact (reach and significance), if these are available.

The contribution of researchers to the public discussion of science is clearly essential. Funding organisations have worked to support this dialogue and promote public engagement with science. There is clearly also interest in the ways in which researchers engage policy makers to see if this information can better support the development of evidence-based policy making and advocacy for research organisations.

[47] *Outputs, Outcomes and Impact of MRC Research* (MRC, 2011) http://www.mrc.ac.uk/consumption/groups/public/documents/content/mrc008191.pdf

The DFG awards the ‘Communicator Award – Science Award of the Donors' Association’. This personal award is given to researchers who have communicated their scientific findings to the public with exceptional success. Apart from that, the DFG offers a funding initiative that promotes the transfer of results of basic research into the public (e.g., teaching models developed with partners from the education sector, cooperatively developed exhibitions with museums).

2.8. Use of Research Infrastructures/Research Resources

Research organisations are interested in strategic decisions about the establishment, funding and evaluation of research infrastructures. These are national or international resources, services or facilities that help support the science base. These infrastructures might include databases, large or small shared facilities, collections of samples, etc. Compared to research programmes they present different challenges to funders with respect to the business case for continued investment.

The European MERIL project[48] is systematically identifying research infrastructures and the ESF also supports an MO Forum[49] looking at issues of good practice associated with research infrastructures. The European Strategy Forum for Research Infrastructure (ESFRI) is charged by the Competiveness Council and the European Commission to develop an evaluation and prioritisation scheme which will distinguish at least between research infrastructures with a pan-European dimension and others which will remain important for regional and/or national needs[50]. However, there is much still to do in order to make available good quality data about the contribution that these research infrastructures make to the science base.

To evaluate the impact of research infrastructures, data is needed on the output that results from the use of such resources. Those that manage these infrastructures usually require that users return information about the results of the work. Such facilities would benefit from operating their own systems for output collection, which address the needs of their sponsor agencies. These output data can then be linked to input data from the facilities on the services provided to users, ideally including the full costs of time and materials.

Just as funders need to encourage more accurate and consistent acknowledgement in research publications, research infrastructures benefit from encouraging their users to acknowledge use of their facility in publications in a standard way. Facilities usually require users to properly acknowledge use of the facility in their formal data

---

[48] The MERIL project (Mapping of the European Research Infrastructure Landscape) aims to achieve this comprehensive inventory of research infrastructures of European relevance and make the information publicly available through an interactive online portal. It is funded by the European Commission under Framework Programme 7 - Contract # 262159 and is being coordinated by the ESF. http://www.esf.org/activities/science-policy/research-infrastructures/meril-mapping-of-the-european-research-infrastructure-landscape.html

[49] ESF Member Organisation Forum on Research Infrastructures http://www.esf.org/activities/mo-fora/research-infrastructures.html

[50] http://ec.europa.eu/research/infrastructures/pdf/esfri_evaluation_report_2011.pdf#view=fit&pagemode=none

access/sharing policy. Publications can then be routinely searched for acknowledgement of particular facilities.

In addition, research funders may be interested not only in the facilities they support, but in the entire range of facilities that their researchers access. This information may be helpful in ensuring that funding agencies are aware of the infrastructure upon which their researchers rely, and in identifying competing or duplicating infrastructure.

If many funding agencies share the data they have on the use of facilities, then this wider view significantly enhances the comparisons and analysis that can be carried out.

3. The collection of output data

Working group members identified a number of processes that are important to examine in sharing experiences about the capture and analysis of research outputs.

3.1. Data from researchers, research organisations or repositories?

Ideally, data should be provided/captured from researchers once and once only, and then shared widely. Research funding and research performing organisations are acutely aware of the need to minimise the administrative burden on the research community. If data are captured routinely by research organisations (for instance, for the purposes of performance management or communication), or provided to repositories (e.g., in the case of publication datasets such as PubMed[51], Scopus[52], Web of Knowledge[53], ISBN[54]), then it is preferable to source this information from these repositories.

As it is unlikely that there will be a single shared approach to collecting output data, the joining up of output data with other information about research (e.g., grant portfolios, publication repositories) and the joining up of output datasets (across research funding organisations and research performing organisations) are pressing issues.

As previously noted, the US STARMETRICS programme aims to establish an information infrastructure in which all those involved in research are represented. Information about all inputs and subsequently data on output will be added to this framework. This approach 'follows' the people in the research system ('actors'); other inputs such as project/programme funding are treated as interventions causing the actors to be more or less productive.

[51] PubMed comprises over 21 million citations for biomedical literature from MEDLINE, life science journals and online books. PubMed citations and abstracts include the fields of biomedicine and health, covering portions of the life sciences, behavioural sciences, chemical sciences and bioengineering. PubMed also provides access to additional relevant web sites and links to the other NCBI molecular biology resources. PubMed is a free resource that is developed and maintained by the National Center for Biotechnology Information (NCBI), at the US National Library of Medicine (NLM), located at the National Institutes of Health (NIH). http://www.ncbi.nlm.nih.gov/pubmed/

[52] Scopus is a proprietary abstract and citation database of research literature and quality web sources covering nearly 18,000 titles from more than 5,000 publishers. Scopus is provided by Elsevier B.V. http://www.scopus.com/home.url

[53] ISI Web of Knowledge is a proprietary set of databases including 23,000 journals, 23 million patent records, and integration with Researcher ID. Web of Knowledge is provided by Thompson Reuters. http://thomsonreuters.com/content/science/pdf/Web_of_Knowledge_factsheet.pdf

[54] The International Standard Book Number (ISBN) is a unique numeric commercial book identifier. Depending on the country of issue, proprietary or free databases are available in which ISBN details can be searched.

This approach requires significant investment in a national information framework to join up data held by universities and funding agencies. Many research funding and research performing agencies have chosen to seek information on output, primarily capturing this directly from researchers or existing repositories, taking the shortest route to capturing information on outputs and impacts. While these data may be linked to information on principle investigators or other management information about staff supported by research funding, data about all the actors in the research process may not be captured. Approaches that largely follow project/programme funding seeking to link outputs to these funding inputs can be taken without the need to establish a new information framework on the whole workforce, but may lack the information needed to fully understand the return on investment.

Ideally, output data systems should be designed to provide the flexibility to analyse output by project or by person.

3.1.1. Frequency/timing of data collection

Information on output is often collected via final grant reports, the advantage being that this is an established part of the grant management process. The disadvantage is that this collects a snapshot of information at the end of the tenure of the grant, prior to some outputs being realised.

Information may be captured on an ongoing basis throughout the lifetime of the grant and beyond. Issues include how long information should be collected after the completion of the award, whether information should be able to be submitted at any time, or annually.

It is also important to consider issues of recall for researchers; if data collection occurs annually are activities that happened earlier in the year less likely to be reported? If researchers are asked to report on output over multiple years, will older outputs be less well represented? This is particularly important when considering how to initially populate datasets; how many years of data will be useful? The issue of recall by researchers is covered in work by RAND Europe[55].

[55] *Strengthening Research Portfolio Evaluation at the Medical Research Council: Developing a survey for the collection of information about research outputs* (RAND Technical Report, 2010) http://www.rand.org/content/dam/rand/pubs/technical_reports/2010/RAND_TR743.pdf (a discussion of researcher recall is on p. 10)

3.1.2. Definition of outputs

Research outputs, as the products generated from research, include the means of evidencing, interpreting and disseminating the findings of a research study.
A holistic approach should be taken to research outputs, seeking to capture output beyond papers and patents (including research materials created, dissemination activities other than publication, etc.). Categorisation of these research outputs is a large topic. Across disciplines there may be many types of output, but it may be possible to determine a manageable core set of output types for areas such as humanities and social sciences, technology and engineering, or biomedicine. With experience it may become apparent that some output types are common across these areas. For example, it has been possible for the UK Medical Research Council and the Science and Technology Facilities Council (STFC)[56] both to collect output data using the e-Val approach, and this experience has demonstrated that, although the research supported by the two research councils is quite distinct, there is the opportunity to standardise definitions in some overlapping output types (e.g., publications, product development and collaboration).

It is worth noting that it is essential to check with the research community to ensure that the data collected appear reasonable, and the questions used to gather them are comprehensible.

3.1.3. Attribution

It is acknowledged that due to the lag times between inputs and impact it is difficult to accurately attribute output. With the STARMETRICS approach the capture of all inputs and all outputs means that attribution, the linkage of outputs to specific inputs/sets of inputs, may not be necessary. In other approaches involving following project and programme investment there may be some general principles regarding attribution:

- Do not apportion output (do not try to measure the relative contribution of different inputs);
- Broadly aim to attribute all output to all funding inputs active in a reasonable timeframe.

3.1.4. Collect what is useful for your evaluation framework

It is important to consider the approaches that will be taken to analyse the data collected and also the eventual strategic questions that may be asked. Capture only

---

[56] The Science and Technology Facilities Council (STFC) supports a broad range of research in particle physics, nuclear physics and astronomy. Information about STFC e-Val can be found at http://www.stfc.ac.uk/About%20STFC/18664.aspx

data that will be used, so that those providing it see the benefit. Ensure good logic models are in place for evaluation programmes, so that collection of each data item can be justified.

3.1.5. Sanctions and incentives for capturing the data

To encourage the submission/collection of output data, research funding and research performing organisations will need to put in place appropriate incentives and sanctions.

Those used by the organisations represented in the working group include:

- Communicating clearly the need for comprehensive, accurate and timely information on the outputs of research funding to the research community and asking for their support;
- Securing support from researchers' employing institutions, such as universities;
- Putting research funding on hold for non-compliant researchers;
- Removing eligibility for research funding from non-compliant researchers;
- Offering services such as the regular provision of output data and appropriate benchmarking information.

Clearly there need to be processes in place for these sanctions to be lifted. For research funding organisations, thought needs to be given to whether sanctions should be replaced by incentives and whether they can be applied at the level of the individual researcher, or their host institution.

3.1.6. CERIF - a standard for the mapping and exchange of data

There is interest in whether the CERIF data model, provided by EuroCRIS, could be a basis for solving the problem of mapping and exchanging data between systems and organisations[57]. CERIF is a European standard and to our knowledge there is no comparable work elsewhere in the world. EuroCRIS has struck collaborations with the Canadian CASRAI[58] and US VIVO[59] initiatives which work on complementary aspects of research information management.

[57] See for example Cerif Tutorial, Lille, 2011, slides 25, 28.

[58] Consortia Advancing Standards in Research Administration Information (CASRAI), a not-for-profit standards development organisation specialising in standardising research data begun in Canadian universities and may be adopted by the main Canadian research funders. CASRAI is developing a common data dictionary for data exchange and reuse between research teams, institutions, and funding agencies throughout the entire life-cycle of research activity http://www.casrai.org/

The working group agreed that output collection systems should work to align their definitions of data with the CERIF standard and provide CERIF compatible output, as the only emerging European standard in this area.

Although further work is needed to extend the CERIF model to encompass all output types covered in this report, it would clearly be helpful to work to this model when designing approaches to exchange and map output data between repositories.

[59] VIVO is an open source ontology and software system designed at Cornell University for researchers and used by some universities in the USA. It is based on the Semantic Web / Linked Open Data concept http://vivoweb.org/

4. Analysis

4.1. Quality Factors

It is important, where possible, to have measures in place to assess the quality of output (e.g., citation impact for papers and information about the licensing of patents). Wherever possible, these 'quality factors' should be derived for output data; research funding and research performing organisations should not rely on volume of output[60]. Counting outputs tends to provide perverse incentives (e.g., 'salami' slicing research findings into a number of papers, when a single paper would have more impact).

A significant barrier to the analysis of output information is the lack of quantitative measures to rank or assess output. The UK REF exercise will assess 'reach' and 'significance' of impacts[61], but these assessments have to be made by peer review panels; there are no rigorous methodologies to apply to the majority of output types. It would be helpful to establish approaches to, even in a semi-quantitative way, 'value' or rate the significance of outputs. Progress in this would clearly make a large difference to the analysis of output data. Approaches to estimate the value of intangible assets might be applied to this problem[62].

4.2. Quality assurance/validation

It is also important to validate the data collected. Outputs should be evidenced/referenced, where possible allowing data to be checked. While much of the feedback from researchers will have to be taken on trust (if it was available elsewhere then researchers would not need to be approached directly), it is important to find ways to assess levels of under- and over-reporting. Cross checking output data against other sources, querying outlying data, and generally 'sense checking' results is advisable.

---

[60] In 2010 the DFG started an initiative to strengthen the focus on quality not quantity of publication output. Since then the number of publications in applications has been restricted to the five most important.

[61] *REF Panel criteria and working methods* (HEFCE, 2012) http://www.ref.ac.uk/media/ref/content/pub/panelcriteriaandworkingmethods/01_12.pdf

[62] For example see *The Value of Intellectual Property, Intangible Assets and Goodwill* (written by Kelvin King, founding partner of Valuation Consulting), published on the WIPO website at http://www.wipo.int/sme/en/documents/value_ip_intangible_assets.htm

4.3. Examples of output collection and analysis

Mapping the Impact: Exploring the payback of arthritis research[63]

This study, carried out in 2007, examined the output from 136 grants awarded by the Arthritis Research Council (now Arthritis Research UK) that had completed research between 2002 and 2006. The report describes the piloting and development of the RAISS online survey tool. The results showed that 2.5% of ARC research awards had led to diagnostics, therapeutics or public health advice that was in or nearing use, and 7.6% had generated intellectual property that had been protected or was in the process of being so.

Analysis and outputs of HRB Grants 2008-2009[64]

This report summarises the outputs from awards supported by the Health Research Board in Ireland that completed research in 2008 and 2009. Output data was compiled from end-of-grant reports submitted four months after the grant terminated. The results identified a wide range of outputs including 8% of awards that had contributed to the generation of new diagnostics or treatments, and 3% that had led to granted patents.

MRC e-Val system (now "Researchfish")[65]

The MRC implemented an online outputs and outcomes collection system in 2008. By 2011 the MRC e-Val dataset had grown to cover all MRC research funded since 2006, and includes a comprehensive range of outputs and outcomes. In 2012 the approach was commercialised, updated and marketed to all funding agencies as a route to collecting output information. A noteworthy development was that the updated system was 'federated'. The advantage of the federated approach is that any number of research organisations can utilise the system to gather output data from researchers, but researchers only have to enter their output data once. Output data can be easily attributed to grants from multiple funding agencies. By October 2012 the system now known as "Researchfish"[66] was being used by 17 research organisations in the UK, which in total spend almost £1.5bn on research each year.

Swiss National Science Foundation (SNF)[67]

In 2011, the SNF began systematically collecting data on the output of funded projects. These data concern, e.g., the education of young scientists, publications, events, patents

---

[63] http://www.rand.org/content/dam/rand/pubs/monographs/2009/RAND_MG862.pdf
[64] http://www.hrb.ie/fileadmin/Staging/Documents/RSF/PEER/Evaluation_docs/2008___2009_Grants_Outputs_report.pdf
[65] http://www.mrc.ac.uk/Achievementsimpact/Outputsoutcomes/e-Val/About/index.htm
[66] www.researchfish.com
[67] http://www.snf.ch/E/current/Dossiers/Pages/output-of-research.aspx

and spin-offs and are inputted by the researchers themselves via the online portal mySNF, with assisting functionalities such as publication import. The SNF intends to use the data for project monitoring and for the evaluation of follow-up projects. At the same time, the data will offer a useful means of showing politicians and the public more tangibly the wide range of research results obtained through the SNF's funding activities. The SNF supports all research disciplines.

German Research Foundation (DFG)

Besides an annual monitoring survey in the coordinated funding programmes, Collaborative Research Centres, Research Training Groups, Clusters of Excellence and Graduate Schools, the DFG collects output information of funded projects in final reports. So far these reports are PDF or paper documents. Since 2010 abstracts and publication lists of these reports are published in GEPRIS - German Project Information System[68] so they are available to the public. The DFG is also planning to implement a web-based system to systematically collect and analyse the results of funded projects. A standardised reporting structure will facilitate the preparation of final reports and systematically compile information that is relevant for evaluation.

Sweden, Prisma

In Sweden, the Swedish Research Council, FAS and Formas are jointly developing Prisma, a system for handling research proposals and collecting output data, drawing upon information from online CVs with assisting functionalities such as publication import. The system is planned to be launched in November 2012.

Norway, CRIStin

CRIStin (Current research information system in Norway[69]) is a research information system for hospitals, research institutes, and universities and university colleges. One of the primary purposes of the system is to collect all the registration and reporting of research activities of institutions within the three sectors in a common system. This gives researchers a place to capture and simplify the registration of common publications. CRIStin as an organisation works closely with the University Centre for Information Technology (USIT) at the University of Oslo, and the CRIStin system has been developed by system developers there.

---

68 www.dfg.de/en/gepris

69 http://www.cristin.no/english/system/

5. Conclusions and Recommendations

The definition, collection, analysis and sharing of output data is a fast moving field. Only three to four years ago there were few systematic, and no cross-funder, approaches to gathering information about research outputs. In 2011/12 the first cross-funder systems for output collection were launched in the UK; in the next few years there may be five to ten systematic large-scale output collection initiatives across Europe alone.

This will highlight the challenges in linking, validating and achieving best value from these data. Research organisations will want to move from a focus on local evaluation of progress, productivity and quality to national and international analysis of impact and more sophisticated benchmarking and comparisons.

The opportunity for member organisations to exchange ideas and to discuss plans within the ESF Forum has been extremely helpful.

We recommend that:

- Member organisations should continue to exchange views and good practice with respect to output collection;
- Research organisations support international efforts to uniquely identify researchers, and link this information to research output information;
- Output collection systems should work to align their definitions of data with the CERIF standard and provide CERIF compatible output;
- Rigorous methodologies for the valuation of output should be identified.

6. Glossary

This glossary is not meant to be exhaustive. Readers are directed to resources such as the OECD's glossary of key terms and concepts[70] which provides extensive definitions of many relevant terms.

Commercialisation

Commercialisation is the process of transforming knowledge into money by trading intellectual property in the market place. Once IP rights have been secured via a national or regional patent office (for example), the main commercialisation routes available to researchers are:

1. Licence to a company;
2. Collaboration with a commercial partner;
3. Start-up or spin-out company.

Current Research Information System (CRIS)

A Current Research Information System, commonly known as 'CRIS', is any informational tool dedicated to provide access to and disseminate research information. A CRIS consists of a data model describing objects of interest to R&D and a tool or set of tools to manage the data.

A CRIS aims to assist the users in their recording, reporting and decision making concerning the research process, whether they are developing programmes, allocating funding, assessing projects, executing projects, generating results, assessing results or transferring technology.

At institutional level it is a tool for policy making, evaluation of research based on outputs, documenting research activities and output and assistance in project planning, and constitutes a formal log of research in progress. For the individual end users a CRIS is essential to evaluate opportunities for research funding, avoid duplication of research activity, analyse trends, have references/links to full text or multimedia scholarly publications, locate new contacts/networks and identify new markets for products of research. Typical forms of output are researcher CV, management information, reports to funders, research bibliography and commercial output reports (Source: EuroCRIS).

[70] http://www.oecd.org/dac/glossary

Intellectual property (IP)
Intellectual property (IP) refers to creations of the mind: inventions, literary and artistic works, and symbols, names, images, and designs used in commerce. IP is divided into two categories: industrial property, which includes inventions (patents), trademarks, industrial designs and geographic indications of source; and copyright, which includes literary and artistic works such as novels, poems and plays, films, musical works, artistic works such as drawings, paintings, photographs and sculptures, and architectural designs. Rights related to copyright include those of performing artists in their performances, producers of phonograms in their recordings, and those of broadcasters in their radio and television programmes (Source: WIPO).

Countries have laws to protect intellectual property for two main reasons. One is to give statutory expression to the moral and economic rights of creators in their creations and the rights of the public in access to those creations. The second is to promote, as a deliberate act of government policy, creativity and the dissemination and application of its results and to encourage fair trading which would contribute to economic and social development.

Patents
A patent is a document issued, upon application, by a government office (or a regional office acting for several countries), which describes an invention and creates a legal situation in which the patented invention can normally only be exploited (manufactured, used, sold, imported) with the authorisation of the owner of the patent. Simply put, a patent is the right granted by the State to an inventor to exclude others from commercially exploiting the invention for a limited period, in return for the disclosure of the invention, so that others may gain the benefit of the invention. Once patents are filed, it typically takes 18 months for the details to be published and once published there is no longer any issue of prior disclosure. The protection conferred by the patent is limited in time (generally 20 years).

Technology Transfer
The term 'technology transfer' is used to describe the process of moving the commercial outputs of a research project out of a higher education institute and into a company or commercial environment.

7. Acknowledgements

This report was compiled by Ian Viney (Chair of the MO Forum working group and Head of Strategic Evaluation, UK Medical Research Council), who gratefully acknowledges the significant input from working group members who contributed to lively and informative discussion in the ESF Member Organisation Forum on Evaluation of Publicly Funded Research over the last four years. In particular, the drafting of this document relied upon the contributions of Brendan Curran (HRB), Jenny Nordquist (Swedish Research Council), Ingrid Roxrud (Research Council of Norway), Katharina Fuß (DFG) and Katrin Milzow (Swiss National Science Foundation).

Annex 1

Organisations responding to the ESF international survey

Organisation	Contact	Country/Region
Agency for Science, Technology and Research (A*STAR)	Tricia Huang	Singapore
Czech Science Foundation (GACR)	Veronika Paleckova	Czech Republic
Danish Agency for Science, Technology and Innovation	Claus Beck-Tange	Denmark
Deutsche Forschungsgemeinschaft (DFG)	Katharina Fuß	Germany
Economic and Social Research Council (ESRC)	Vicki Crossley	United Kingdom
Engineering and Physical Sciences Research Council (EPSRC)	Sue Smart	United Kingdom
European Science Foundation (ESF)	Farzam Ranjbaran	Europe
European Commission	Peter Fisch	Europe
Fonds de la Recherche Scientifique (FNRS*)*	Nadège Ricaud	Belgium
Formas	Bengt Ohlsson	Sweden
Foundation for Polish Science	Marta Lazarowicz-Kowalik	Poland
Health Research Board of Ireland	Brendan Curran	Ireland
Hungarian Scientific Research Fund (OTKA)	Gyula P. Szigeti	Hungary
Istituto Nazionale di Fisica Nucleare (INFN)	Valerio Vercesi	Italy
Institut national de la santé et de la recherche médicale (INSERM)	Isabelle Henry	France
Instituto de Salud Carlos III	Rafael De Andrés Medina	Spain
Medical Research Council	Ian Viney	United Kingdom
Ministerio de Ciencia, Tecnología e Innovación Productiva (MINCYT)	Agueda Menvielle	Argentina
National Health and Medical Research Council (NHMRC)	Marcus Nicol	Australia
National Research Council of Italy (CNR)	Massimiliano Di Bitetto	Italy
Natural Environment Research Council (NERC)	Gregor McDonagh	United Kingdom
Netherlands Organisation for Scientific Research *(*NWO*)*	Anko Wiegel	Netherlands
Research Council of Norway	Ingrid Roxrud	Norway
Science and Technology Facilities Council (STFC)	Isobel Climas	United Kingdom
Science Foundation Ireland (SFI)	Helen O'Connor	Ireland
Slovak Academy of Sciences	Iveta Hermanovská	Slovakia
Spanish National Research Council (CSIC)	Juan de Damborenea	Spain
Swedish Research Council	Jenny Nordquist	Sweden
Swiss National Science Foundation	Katrin Milzow	Switzerland
The Scientific and Technological Research Council of Turkey (TUBITAK)	Mustafa Ay	Turkey