Towards Data Integration for Research Funding and Performing Organisations: a Science Europe Initiative

Science Europe[1]

Submission to:
'OECD Blue Sky Forum on Science and Innovation Indicators – Informing Science and Innovation Policies: Towards the Next Generation of Data and Indicators'
Ghent, 19-21 September 2016

Abstract

Research funding and performing organisations collect and use data from various and heterogeneous sources for different purposes, such as monitoring and evaluating research activities and outputs, funding allocation, strategic decision-making, and describing their systemic role to policy-makers, stakeholders and the public.

Such data – referred to in this article interchangeably as 'data on research activity' and 'research information' – provide information at aggregate and individual level for example on the careers of researchers, on the success rates of calls for research proposals, on the evaluation of such proposals, on the evolution of research disciplines and budgets, on research performance (via bibliometric indicators), on the collaboration between research institutions, and on the funding streams that contributed to individual outputs.

To a large extent, the data on research activity collected and compiled by research organisations is not interoperable, which means that it is difficult to integrate data from different sources and to reuse such data in different contexts, for example for cross-organisational and cross-country comparisons or for studies and analyses on science, technology and innovation.

---

[1] Science Europe (SE) is an association of 47 Research Funding and Performing Organisations with a public mission and a substantial impact on their national research system. Together, members invest about €25 billion in research and development activities per annum. A full list of SE members is available here: http://www.scienceeurope.org/about-us/member-organisations.
This text was prepared by the SE Office to disseminate SE activity and does not represent an official position of the organisation, nor should it be portrayed as such.
Many individuals have contributed to this work in different ways, for example by carrying out a survey, by drafting sections of reports and background documents, by contributing to discussions and providing feedback on documents and by organising and managing the work. For the SE Working Group Research Policy and Programme Evaluation (in alphabetical order): Raphaël Beck (F.R.S. – FNRS, Belgium), Emily Gale (MRC, UK), Wolfgang Glänzel (FWO, Belgium), Michal Kołodziejski (NCN, Poland), Katrin Milzow (SNSF. Switzerland), Jordi Molas-Gallart (CSIC, Spain), Lisa Murphy (SFI, Ireland), Stojan Pečlin (ARRS, Slovenia), Matteo Razzanelli (Science Europe Office), Pei-Shan Chi (KU Leuven, Belgium), Stig Slipersæter (RCN, Norway), Gábor Toth (OTKA, Hungary), Bérangère Virlon (ANR, France). In addition, the Science Europe Office wishes to thank the participants to the Science Europe Workshop "Data on Research Activity: Towards Data Interoperability for Research Funding and Research Performing Organisations" (The Hague, 15 and 16 June 2016), and in particular the moderators and keynote speakers: Paul Wouters (CWTS, The Netherlands), Cameron Neylon (Curtin University, Australia), Thed van Leeuwen (CWTS, The Netherlands), Vasco Vaz (FCT, Portugal), Christian Gutknecht (SNSF, Switzerland), Vivien Petras (Humboldt University, Germany), Torger Moeller (DZHW, Germany), Gunnar Siversten (NIFU, Norway). Responsibility for this text and any mistake therein remains with the SE Office.
For correspondence: Matteo Razzanelli, Science Europe Office, Rue de la Science 14, 1040 Brussels (Belgium).
E-mail: office@scienceeurope.org

Science Europe (SE), an association of 47 research funding and performing organisations with a public mission, is engaged, amongst other actions, in strengthening the European research system, and in promoting the standardisation, harmonisation and integration of data on the research activities its members fund or manage. Section 3 of this paper presents a draft SE position statement on research information systems prepared by the SE Working Group Research Policy and Programme Evaluation in order to gather further input and feedback from stakeholders, policy makers and experts.

The idea behind the document is to facilitate progress on research information interoperability by setting out a long-term vision on the use of research information systems, in order to provide Member Organisations (MOs) with a focal point for convergence when they are making decisions, or providing guidance or advice on research information systems in their countries. If this high-level vision is agreed upon by SE, follow-up actions can be undertaken by members, both collectively and individually, as well as by other research organisations wishing to take steps towards enhanced interoperability.

Introduction

Research funding and performing organisations use data from various and heterogeneous sources for different purposes, such as monitoring and evaluating research activities and outputs, funding allocation, supporting decision-making on their policies and strategies, and describing their systemic role to policy-makers, stakeholders and the public.

These data – referred to in this article interchangeably as 'data on research activity' and 'research information' – comprise datasets created and stored by a variety of organisations on entities such as researchers and their CVs, research institutions, funding applications, research projects, research outputs (publications, research data, patents and so on).

The sources of these data include administrative and financial documents (for example funding applications and financial reports), management documents (for example project reports or evaluations carried out to select funding applications), bibliometric databases (mainly Elsevier's Scopus, Web of Science, Google Scholar), surveys, and others.

Such data provide information at aggregate and individual level for example on the careers of researchers, on the success rates of calls for research proposals, on the evaluation of such proposals, on the evolution of research disciplines and budgets, on research performance (via bibliometric indicators), on the collaboration between research institutions, and on the funding streams that contributed to individual outputs..

A SE report[2] finds that research information is largely neither standardised nor interoperable between the major research funding and research performing organisations that make up SE's membership. A workshop held by SE in June 2016[3] shed light on the fact that the situation seems to be similar among other research organisations such as universities across Europe.

[2] Science Europe, 2016
[3] See section 2 below.

As a result, integrating research information, reusing it in different contexts (for example for evaluation studies or to monitor policy compliance), and comparing research activities across organisations and institutions is difficult and costly. Moreover, the inability to reuse data across institutions and organisations engenders the need for research performers to provide the same data multiple times thus increasing the cost of data curation, while decreasing data accuracy. Finally, the lack of standards for the curation and use of research activity data raises questions about their quality, reliability and appropriate use for evaluation purposes.[4]

Interest in data on research activity goes well beyond research funding and performing organisations: decision-makers in government and elsewhere, individual researchers – especially those engaged in the 'science of science policy'[5] and other social scientists – also depend directly or indirectly on research activity data, their mutual compatibility and interoperability.

Among other areas of activity, SE is engaged in promoting the standardisation, harmonisation and integration of data on the research activities funded or managed by research organisations with a public mission. In particular, the strategic plan of SE, its 'Roadmap'[6], sets the over-arching objective of contributing *"to advancing the development and implementation of standards for the definition, acquisition, storage, analysis and sharing of evaluation-related data".*[7]

In brief, one of SE's aims is to progress research information standardisation and interoperability.

This paper reports on the outcomes of the efforts led by SE in this context, based on the results of a workshop organised by SE on 15-16 June 2016 in The Hague, entitled 'Data on Research Activity: Towards Data Interoperability for Research Funding and Research Performing Organisations'. The workshop was attended by representatives of SE Member Organisations (MOs), as well as by external stakeholders and experts. As a basis for discussion, workshop participants used an Interim Report produced by SE by the same title[8]. The report contains the outcomes of a survey conducted by SE to map the practices of MOs in terms of the collection and use of research activity data.

Even if SE focuses on research information interoperability in the context of its membership, it is clear that the interoperability of research information involves collaboration and interaction among all the actors of a complex eco-system of organisations that, beyond SE's membership, includes universities, scientific publishers and their journal databases, the academic community (for example research policy scholars and bibliometricians), research policy consultants, information infrastructures such as national Current Research Information Systems, national

[4] Glänzel (1996), Wouters et al. (2013), and Hicks et al. (2015).
[5] According to van den Besselaar et al. (2016): "Up to now, STI studies are either *rich* but small scale (qualitative case studies) or large scale and *under-complex* – because they generally use a single dataset like Patstat, Scopus, WoS, OECD STI indicators, etc., and therefore deploying only a few variables – determined by the data available. However, progress in the STI research field depends in our view on the ability to do large-scale studies with often many variables specified by relevant theories: There is a need for studies which are at the same time big *and* rich."
[6] December 2013, available at: http://www.scienceeurope.org/uploads/PublicDocumentsAndSpeeches/ScienceEurope_Roadmap.pdf
[7] Source: Science Europe Roadmap, p. 27.
[8] Science Europe, 2016, available at: http://scieur.org/data-stand

and international publication repositories such as OpenAIRE[9], standardisation organisations (such as euroCRIS[10], CASRAI[11], ORCID, CrossRef[12], DataCite[13]), information systems suppliers and policy makers. The June 2016 SE workshop was used to involve representatives from several of these communities in the discussion, in order to help SE focus on the issues at stake and scope realistic approaches to help meet its data interoperability objectives.

The outcome was a draft SE Position Statement, prepared after the workshop by the members of the SE Working Group Research Policy and Programme Evaluation who participated, and further improved with the comments of workshop participants, who collaborated remotely. The draft statement formulates a proposal for a SE long-term vision for the interoperability of research activity data by focusing on the information system storing them.

The document lays out four fundamental principles for research organisations to consider when making decisions on their research information systems and, based on those principles, lists a series of concrete follow-up actions that can be undertaken by undertaken by SE members, both collectively and individually, to achieve real progress in terms of data interoperability.

The potential impact this document could have is to become a reference for research organisations that need to grapple with issues related to the interoperability of research activity data, and to pave the way for cross-institutional collaboration in this respect.

It is important to note that the workshop outcome document presented below (section 3) is not an official SE Position Statement, nor has it yet been submitted for approval by MOs. The objective of this paper is to use the opportunity provided by the OECD Blue Sky III to elicit feedback from relevant stakeholders and experts, in order to further improve this first draft ahead of its approval process.

This paper is structured as follows:

- Section 1 traces the origins and context of SE's engagement with the interoperability of research activity data, and provides an overview of the current state of affairs among SE MOs on this respect;
- Section 2 briefly describes the structure and some of the main results of the workshop that led to the draft SE position statement; and
- Section 3 reproduces the latest draft of the position statement prepared by the SE Working Group Research Policy and Programme Evaluation with input from workshop participants.

[9] https://www.openaire.eu/
[10] www.eurocris.org/
[11] http://casrai.org/
[12] www.crossref.org/
[13] https://www.datacite.org/

1. Science Europe Engagement with the Interoperability of Research Activity Data[14]

Origins and Context

Discussions between SE MOs on the interoperability of research information date back to SE's predecessor organisation, the European Science Foundation (ESF) [15], who organised the collaboration of MOs on *ex-post* evaluation issues dating back to 2007, under the aegis of a few different 'ESF Fora'.

Those Fora had concluded that the *"standardisation of data collection is an ambitious but much needed objective"*,[16] and suggested pressing forward on the interoperability of classification systems, the use of shared unique identifiers for researchers, grants and institutions (in order to open up the possibility to link different datasets held in different databases), as well as the adoption of common standards for information systems, such as CERIF[17].

In addition, potential was identified in the harmonisation and standardisation of evaluation-related rules and processes, such as policies on the acknowledgement of funding sources in papers, the disclosure of information related to evaluations, and the processes used for bibliometric analysis.

The ESF identified three overarching rationales for taking actions in the areas listed above[18]. Firstly, because datasets can be used in different evaluation contexts to produce tailor-made indicators and feed as evidence into tailor-made evaluation studies. Secondly, because data collection presents a clear burden to applicants and grantees, hence its collection should be minimised, and the data should only be collected once and re-used in different contexts. Thirdly, because MOs should increase the comparability of their activities so that they can better "position themselves on the European and global research landscape"[19].

It should be noted that as the reflections and analyses described in this section went on, in the period 2007-2016, the research information landscape continued to evolve rapidly both at national and international level, and important organisations and initiatives have emerged or have further developed in the meantime.

Many countries created or expanded their national Current Research Information Systems (CRIS)[20], often based on the CERIF standard data model[21]. In Germany, a major project called Research Core Dataset[22] produced a single data model based on CERIF that all German research institutions are invited to adopt.

---

[14] This section reproduces parts of the introduction and executive summary of the Interim Report (Science Europe, 2016).
[15] http://www.esf.org/home.html
[16] European Science Foundation (2012), p. 11.
[17] http://www.eurocris.org/cerif/main-features-cerif
[18] European Science Foundation (2012).
[19] European Science Foundation (2012), p. 14.
[20] For the examples of Norway's CRIStin and Slovenia's COBISS, see the Interim Report, pp. 21-22. See also Mahieu et al. (2014), p. 48.
[21] http://www.eurocris.org/cerif/main-features-cerif
[22] See Science Europe (2016), p. 22 and http://www.forschungsinfo.de/kerndatensatz/en/index.php?home

A large set of standardisation organisations[23] have worked on the development or promotion of interoperability standards, for example with regards to data models, definitions, unique identifiers for entities such as researchers, research grants or institutions.

In addition, a few notable initiatives have focused on process standardisation, by providing guidance on data collection, curation and use, and evaluation:

- Austrian 'Standards of Evaluation in Research and Technology Policy'[24]: guidance issued by the Austrian Platform for Research and Technology Policy Evaluation (fteval)[25] on minimum quality requirements for evaluation processes, that help individuals and organisations create a common evaluation language and culture and avoid potential mistakes;
- Leiden Manifesto for Research Metrics[26]: an initiative by a group of scholars aiming to raise awareness on malpractice regarding the use of metrics in evaluative contexts. The manifesto enumerates principles such as 'avoid misplaced concreteness and false precision' and 'account for variation by field in publication and citation practices';
- 'The Metric Tide' Report'[27]: the 2015 report is the final output of an independent review of the current and future roles of quantitative indicators in the assessment and management of research. The review was commissioned to James Wilsdon by the Higher Education Funding Council for England (HEFCE)[28]. The report proposes the notion of 'responsible metrics' *"as a way of framing appropriate uses of quantitative indicators in the governance, management and assessment of research"*[29]. The report holds that responsible metrics are robust, humble, transparent, diverse and reflexive.[30] In addition, the report contains specific recommendations on data infrastructures, encouraging the use of standard unique identifiers and the linkage of datasets held by different organisations.[31]

Linking the SE discussions to these parallel initiatives and developments was therefore a key requirement for the June workshop described in Section 2.

[23] These organisations notably include EuroCRIS, CASRAI, ORCID, CrossRef, and DataCite.
[24] Available at: http://fteval.at/upload/fteval_Standards_english.pdf
[25] "Founded in 1996 as an informal cooperation, the Austrian Platform for Research & Technology Policy Evaluation aims at presenting approaches and methods of evaluation, discussing the current evaluation practice on an international level and thus contributing to the development of a culture of evaluation in Austria. In November 2006, its members re-founded the Platform Research & Technology Policy Evaluation as a society." Source: http://fteval.at/en/platform
[26] Hicks et al. (2015).
[27] Wilsdon et al. (2015).
[28] http://www.hefce.ac.uk/
[29] Wilsdon et al. (2015), pp. 134.
[30] Wilsdon et al. (2015), pp. 134-135.
[31] Wilsdon et al. (2015), pp. 144-145.

Science Europe Survey of Member Organisations

Discussions within SE in the course of 2014[32] led to the conclusion that MOs still regarded the areas identified by the ESF as relevant, with the only exception of convergence towards a single standard for research information systems.

As a result, SE set out to update the work done by the ESF, by launching an internal survey to map MO practices in relation to data collection and use[33]. The outcomes of the survey are described in detail in the Interim Report[34], which summarises the key learnings from the survey as follows:

- MOs collect and publish different datasets: no clear patterns emerge in terms of the datasets made available and the data sources used to build research activity datasets. However, most MOs are considering some kind of standardisation of the data they collect;
- Most MOs do not use indicators, and bibliometric indicators are by far the most widely used (number of publications, indicators based on journal impact or on the number of citations received);
- Since bibliometric indicators are the most widespread type of indicator in use, bibliographic databases play a key role in research evaluation. To source bibliometric indicators, the most widely used databases are Elsevier's Scopus, Thomson Reuters' Web of Science and Google Scholar;
- Most MOs use a classification system to categorise research activities or outputs by field or discipline. Most MOs use their own, tailor-made classification system, or a national system. In general, conversion tables to make different systems interoperable are not in use. The diversity of classification systems is not seen as a problem;
- Unique researcher identifiers are mostly in use among research funding organisations, but they do not seem to be widespread among research performing organisations. In most cases, those organisations that use unique researcher identifiers opt for an internal system, rather than a standard one. This means that identifiers are not widely used to link data across databases. However, a good number of organisations are adopting or considering the adoption of the global, standardised system called ORCID;
- MOs do not seem to be mining publications to look for funding acknowledgements, suggesting that funding acknowledgement clauses so far have been of limited use for the availability of data linking funding with research outputs.

The picture that emerged from the survey is therefore rather fragmented, and MOs found it difficult to converge on shared recommendations based on the results of the survey alone.

As the relevant MO representatives tried to scope recommendations for short-term actions and longer-term roadmaps, it became clear that it would not be possible to converge on

[32] In particular, such discussions were held by the SE Working Group Research Policy and Programme Evaluation between 2014 and 2015, in order to decide on the activities to include in its 2-years work plan (2015-2017).
[33] Seven areas were covered: classifications of fields of science and technology, data on research outputs, researcher identification, indicators for evaluation purposes, data standardisation, use of bibliographic databases, and acknowledgement of funding sources in scientific publications. See Science Europe 2016 for more details.
[34] Science Europe (2016), p. 5.

shared solutions without an in-depth discussion of the rationales for standardisation and its objectives.

The programme of the previously mentioned June SE workshop was therefore structured so that the identification of shared principles for research information and of real-world use-cases among MOs for data on research activity would be discussed prior to the development of specific technical solutions.

2. Science Europe Workshop 'Data on Research Activity: Towards Data Interoperability for Research Funding and Research Performing Organisations' (The Hague, 15-16 June 2016)

The workshop comprised six discussion themes. The first two themes were about identifying shared principles among MOs for data on research activity (Theme 1), and relevant, real world, use scenarios for such data (Theme 2).

The results of these first two sessions were used in the following ones to discuss and select appropriate technical solutions to implement the principles identified, and address the use scenarios scoped as part of Theme 2. The areas for technical solutions to be discussed were chosen among those scoped by the ESF and the MO survey, and they were respectively:

- researcher identification;
- subject classification systems;
- funding sources identification for scientific publications; and
- combining datasets in evaluative contexts.

With regards to use scenarios that require the interoperability of research activity data across organisations and countries, three broad categories of use cases, from the point of view of MOs, emerged from the discussion:

- Cases in which data quality and completeness depends on interoperability:
  - Linkage of outputs to funding sources;
  - Coverage of diverse outputs beyond publications, such as research datasets; and
  - Coverage of any other kind of information that cannot be internally sourced.
- Strategy and planning:
  - For research performing organisations, insights on collaborations and funding opportunities;
  - Information on co-funding with other European agencies;
  - Comparisons of organisational and demographic performance (including on questions of diversity);
  - Career tracking; and
  - Disciplinary analysis and comparisons across countries.
- Policy development and implementation:
  - Cross-checking for compliance with policy (for example: implementation of Open Access policies); and
  - Intelligence on salient issues, for example article processing charges data in negotiations with publishers.

In addition, workshop speakers shared their experience with research information in universities and in standards organisations such as EuroCRIS. While standardisation has progressed and is further being developed by standards organisations, consortia and national infrastructures and initiatives, the lack of interoperability of research information across organisations and countries goes much beyond SE's membership, and also concerns universities and data analytics services offered by private service providers such as scientific publishers.

The MO use cases for interoperable data on research activity listed above, together with the experiences shared by other data stakeholders at the workshop, constitute the backdrop against which a set of shared principles and potential actions was scoped during the discussions and consolidated by the SE Working Group Research Policy and Programme Evaluation immediately after the workshop.

The idea behind the document reproduced in the next section is to facilitate progress on research information interoperability by setting out a long-term vision on the use of research activity data, in order to provide research organisations with a focal point for convergence when they are making decisions, or providing guidance or advice on research information systems in their countries. If such a high level vision is agreed upon by SE, follow-up actions can be undertaken by members, both collectively and individually, as well as by other research organisations wishing to take steps towards improved interoperability.

3. Draft Science Europe Position Statement Open to Internal and External Input and Feedback[35]

The way knowledge is created, shared and applied constantly evolves. New ways to organise research processes are enabled by technological development. Societal values and expectations towards publicly-funded research also evolve, as do the societal challenges addressed by science. The Open Science agenda captures the latest trends in terms of organisation of the scientific enterprise and the societal expectations of it. Open Science is an example of how quickly the context, needs, and objectives related to research systems can evolve. The pace, direction and nature of such changes is unpredictable.

This calls for research information systems that are capable of effectively supporting this constant and unpredictable change with the intelligence and insights needed to perform strategic, analytical and management functions, as well as in order to fulfil their public missions.

Research information systems store data on research activity, such as on scientific publications and other outputs, researchers, research budgets and projects, research institutions, research funding applications and reviews.

Through aggregation, such data provide crucial information for strategic decision making and for science, technology and innovation analyses and studies, for example on the careers of researchers, on the success rates of calls for research proposals, on the evaluation of such proposals, on the evolution of research subjects and budgets, on research performance (via bibliometric indicators), on the collaboration between research institutions, and on the funding streams that contributed to individual outputs..

Science Europe therefore invites all research organisations, including its members, to develop resilient research information systems by adopting the following core principles:

1. **Flexibility:** research information systems should be flexible enough to allow for extensions in terms of the data objects covered, their definitions, metadata, and use of external data sources;
2. **Openness:** research information systems data should be available for external use – in line with the principle '*as open as possible, as closed as necessary*' – and their processing should never require the loss of ownership in underlying raw data by the originating institution;
3. **FAIRness:** research information systems should foster the findability, accessibility, interoperability, and reusability of the data they store by implementing the FAIR Guiding Principles[36] for research activity data;
4. **Data entry minimisation:** research information systems should minimise the need for entering data and facilitate the reuse of data entered manually, in line with the motto '*enter once, reuse multiple times*'.

[35] This draft statement can be downloaded at : http://scieur.org/oecd-poster

[36] Wilkinson, M. D. et al. (2016), *The FAIR Guiding Principles for scientific data management and stewardship*, Scientific Data 3, 15 March 2016, doi:10.1038/sdata.2016.18, retrieved online on 21 June 2016 from http://www.nature.com/articles/sdata201618. The list of principles is listed in the Annex to this document.

These four principles should always be implemented in light of the applicable legal and ethical standards relevant for data handling where the research information system is located. Limitations to the application of these principles may arise due to privacy protection, security and other legitimate concerns. Such limitations should always be applied only insofar as needed to address a valid concern.

Science Europe Member Organisations wish to set an example for the implementation of these principles. For this reason, they have identified a set of four follow-up actions where it is possible to make progress towards their implementation in a concerted way:

1. Combining Data

- Strive to make data on publicly-funded research activity publicly available, ensuring that there are no legal, confidentiality, intellectual property or privacy issues;
- Promote the adoption of research information systems with the following characteristics:
    - Economic value-creation for service providers that does not depend on siloing and closing data;
    - No pay-walling of research activity data, given that dataset combination is a basic requirement to understand the research landscape;
    - Interoperability based on a common exchange format and open standards for the definitions of the entities (such as researchers, organisations, grants, activities, and outputs) and attributes in the research information domain, and for the identification of these entities with unique and persistent identifiers.
- Support efforts to develop infrastructures that enable and facilitate the connection of datasets (including open standards such as CERIF and CASRAI), and other policy initiatives[37] that move in a similar direction to what is advocated here.

2. Funder and Grant Identification

- Foster the compliance of researchers with funding acknowledgement policies by:
    - Co-ordinating, harmonising and issuing a standardised format of preferred acknowledgement text;
    - Developing and co-ordinating on an intelligent standard format for grant IDs and providing a shared database and metadata for them;
    - Engaging with CrossRef to complete and enrich the Funding data[38];

[37] For example EU initiatives such as ESFRI (http://www.esfri.eu/), the European Open Science Cloud (http://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud), RISIS (http://risis.eu).

[38] CrossRef Open Funder Registry (formerly FundRef, http://www.crossref.org/fundingdata/registry.html) is an open registry that provides taxonomy of currently 13000 standardized funder names. Publishers or their manuscript tracking system vendors incorporate the Open Funder Registry into the submission processes. Authors select funders from this list and provide grant numbers at the time of manuscript submission. Publishers send funder information (funder names, funder IDs and grant numbers) to Crossref as part of their regular metadata deposits. Funders can query Crossref and receive DOIs and metadata for articles resulting from their funding (see: http://search.crossref.org/funding).

- Engaging with publishers[39] to use the CrossRef Open Funder Registry during the article submission process and provide funding information via CrossRef or DataCite metadata.

3. Researcher Identification

- Adopt global unique identifiers for researcher identification to support interoperability. By virtue of its open, non-proprietary and independent nature, ORCID[40] is deemed to be the most promising initiative;
- Engage with ORCID to better identify use cases and current issues on data quality and routes to improvement.

4. Subject Classifications

- Provide full documentation on classification systems, including subject definitions and categorisation methodologies, in order to:
  - Support comparisons across datasets held by different organisations, for instance in order to compare data on different research disciplines;
  - Encourage further discussion on harmonisation of and cross-mapping between different types of classification systems.

Annex to the Draft Position Statement - The FAIR Guiding Principles for scientific data management and stewardship

The FAIR principles provide a guideline for those wishing to enhance the reusability of their data holdings: these principles put specific emphasis on enhancing the ability of machines to automatically find and use the data, in addition to supporting its reuse by individuals.

To be Findable:
F1. (meta)data are assigned a globally unique and persistent identifier
F2. data are described with rich metadata (defined by R1 below)
F3. metadata clearly and explicitly include the identifier of the data it describes
F4. (meta)data are registered or indexed in a searchable resource

To be Accessible:
A1. (meta)data are retrievable by their identifier using a standardized communications protocol
A1.1 the protocol is open, free, and universally implementable
A1.2 the protocol allows for an authentication and authorization procedure, where necessary
A2. metadata are accessible, even when the data are no longer available

To be Interoperable:
I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
I2. (meta)data use vocabularies that follow FAIR principles
I3. (meta)data include qualified references to other (meta)data

[39] See publishers that currently provide standardized funding information via CrossRef: http://www.crossref.org/06members/fundrefdeposits.html
[40] ORCID provides a persistent digital identifier that is unique to each researcher and, through integration in key research workflows such as manuscript and grant submission, supports automated linkages between different databases. Source: orcid.org.

I4. (meta)data can be exchanged through a standard format

To be Reusable:

R1. meta(data) are richly described with a plurality of accurate and relevant attributes
R1.1. (meta)data are released with a clear and accessible data usage license
R1.2. (meta)data are associated with detailed provenance
R1.3. (meta)data meet domain-relevant community standards

Source: List of FAIR principles retrieved from Wilkinson, M. D. et al. (2016), *The FAIR Guiding Principles for scientific data management and stewardship*, Scientific Data 3, 15 March 2016, doi:10.1038/sdata.2016.18, retrieved online on 21 June 2016 from http://www.nature.com/articles/sdata201618.https://www.force11.org/group/fairgroup (retrieved on 23/06/2016); introduction to the annex adapted from https://www.force11.org/group/fairgroup (retrieved on 23 June 2016).

Conclusions

The draft SE position statement in Section 3 is a basis for consultation: inputs from MOs, experts and stakeholders will be used to further improve the draft which will be considered by the dedicated SE Working Group on Research Policy and Programme Evaluation and the SE General Assembly.

References

European Science Foundation (2012), *Evaluation in Research and Research Funding Organisations: European Practices – A report by the ESF Member Organisation Forum on Evaluation of Publicly Funded Research*. European Science Foundation: Strasbourg.

Glänzel, W. (1996), *The need for standards in bibliometric research and technology.* Scientometrics, 35(2), 167-176.

Hicks, D., Wouters, P., Waltman, L., de Rijcke, S., Rafols, I. (2015), *Bibliometrics: The Leiden Manifesto for research metrics*. Nature, 520, 23 April 2015, 429–431.

Mahieu, B., Arnold, E. and Kolarz, P. (2014), *Measuring scientific performance for improved policy making*. Science and Technology Options Assessment (STOA) Study, European Parliamentary Research Service (EPRS), European Union: Brussels

Bilder, G., Lin, J., Neylon, C. (2015): *Principles for Open Scholarly Infrastructure-v1*, http://dx.doi.org/10.6084/m9.figshare.1314859 Retrieved: 10:54, May 20, 2016 (GMT)

Science Europe (2016), *Science Europe Interim Report - Data in Research Activity: Towards Data Interoperability for Research Funding and Performing Organisations*, Science Europe: Brussels, available at: http://scieur.org/data-stand

van den Besselaar, P. Khalili A., Idrissou, A., Loizou, A., Schlobach, S., van Harmelen, F. (2016), *SMS: a linked open data infrastructure for science and innovation studies*, Proceedings of the 21st International Conference on Science and Technology Indicators, Valencia (Spain), 14-16 September 2016, forthcoming

Wilkinson, M. D. et al. (2016), *The FAIR Guiding Principles for scientific data management and stewardship*, Scientific Data 3, 15 March 2016, doi:10.1038/sdata.2016.18, retrieved online on 21 June 2016

Wilsdon, J., et al. (2015), *The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management*. DOI: 10.13140/RG.2.1.4929.1363

Wouters, P., Glänzel, W., Gläser, J., Rafols, I. (2013), *The dilemmas of performance indicators of individual researchers – an urgent debate in bibliometrics*. ISSI Newsletter, 9(3), 48-53.