Science Europe Position Statement

On Research Information Systems

November 2016

SCIENCE EUROPE
Shaping the future of research

November 2016

'Position Statement on Research Information Systems': D/2016/13.324/11

Author: Science Europe
Co-ordination: Science Europe Working Group on Research Policy and Programme Evaluation

For further information please contact office@scienceeurope.org

Introduction

Research Performing Organisations (RPOs) and Research Funding Organisations (RFOs) collect and use data about their own activities from various and heterogeneous sources. This kind of data – data about research activities rather than research data generated by researchers – is stored in research information systems.

RPOs and RFOs use research information systems for a variety of different purposes, such as monitoring and evaluating research activities and outputs, allocating funding, supporting decision making on their policies and strategies, tracking researchers' careers, and describing their systemic role to policy-makers, stakeholders and the public.

As a result, decision makers and research organisation managers alike increasingly depend on indicators, reports and studies that draw data from research information systems. Powerful analyses can be derived when different datasets are linked. The reliability, reach and comprehensiveness of these analyses is a function of the quality of the underlying data, their mutual compatibility and their interoperability.

However, this is no easy task, since at present the datasets of research organisations differ widely, for instance with regards to definitions, classification systems and formats. These differences are due to differing cognitive, organisational and administrative needs and standards.

This Science Europe Position Statement is a contribution to enhancing the interoperability of research information systems by providing a set of common principles to guide their development.

Research Information Systems

The way in which 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 the organisation of 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 are 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.

Research information systems store data on research activity, such as 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 to develop resilient research information systems by adopting the following core principles:

Core principles	
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.
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 EU Directive 2013/37/EU[1] – and their processing should never require the loss of ownership in underlying raw data by the originating institution.
FAIRness	Research information systems should foster the findability, accessibility, interoperability, and reusability of the data that they store by implementing the FAIR Guiding Principles[2] for research activity data.
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'.

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 core principles. For this reason, they have identified a set of four follow-up actions through which 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; and
  - 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[3] and CASRAI[4]), and other policy initiatives[5] 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 with respect to, an intelligent standard format for grant IDs, and providing a shared database and metadata for these;
  - Engaging with CrossRef to complete and enrich the Funding data[6]; and
  - Engaging with publishers[7] 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[8] is deemed to be the most promising initiative.
- Engage with ORCID to better identify cases of use 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; and
  - Encourage further discussion on harmonisation of, and cross-mapping between, different types of classification system.

Annex

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 in the table)
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
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 (retrieved on 23/06/2016); introduction to the annex adapted from https://www.force11.org/group/fairgroup (retrieved on 23 June 2016).

Notes and References

1. Directive 2013/37/EU of the European Parliament and of the Council of 26 June 2013 on the re-use of public sector information Text with EEA relevance http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32013L0037
2. 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.
3. Maintenance of CERIF is carried out by euroCRIS, a not-for-profit association that brings together experts on research information in general and research information systems (CRIS) in particular http://www.eurocris.org/cerif/main-features-cerif
4. The Consortia Advancing Standards in Research Administration Information (CASRAI) https://casrai.org/
5. 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), and RISIS (http://risis.eu).
6. 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).
7. See publishers that currently provide standardised funding information via CrossRef: http://www.crossref.org/06members/fundrefdeposits.html
8. 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 http://orcid.org/

Bibliography

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*, https://figshare.com/articles/Principles_for_Open_Scholarly_Infrastructures_v1/1314859 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

Science Europe (2016), *Towards Data Integration for Research Funding and Performing Organisations: a Science Europe Initiative*, 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. Available online at: https://www.scribd.com/document/324194143/056-Towards-Data-Integration-for-Research-Funding-and-Performing-Organisations-a-Science-Europe-Initiative

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.

Science Europe is a non-profit organisation based in Brussels representing major Research Funding and Research Performing Organisations across Europe.

More information on its mission and activities is provided at www.scienceeurope.org.

To contact Science Europe, e-mail office@scienceeurope.org.

SCIENCE EUROPE
Shaping the future of research

Science Europe
Rue de la Science 14
1040 Brussels
Belgium

Tel +32 (0)2 226 03 00
Fax +32 (0)2 226 03 01
office@scienceeurope.org
www.scienceeurope.org