Elsevier values its multi-faceted and synergistic relationship with the NIH and appreciates the opportunity to provide a response to NOT-OD-16-133, Request for Information (RFI): Metrics to Assess Value of Biomedical Digital Repositories. Submitted on behalf of Elsevier by:

Holly J Falk-Krzesinski, PhD
Vice President, Strategic Alliances
Global Academic Relations
Elsevier
h.falk-krzesinski@elsevier.com
New York, NY, USA

Response Contents

Part 1: Research Data Definition and Research Data Metrics

Definition and Disciplinarity

Research Data Definition

Elsevier's working definition is, "research data refers to the results of observations or experimentation that validate research findings." Research data can also be defined as, "the recorded factual material commonly accepted in the scientific community as necessary to validate research findings."[1] Research data covers a broad range of information types[2], and digital data can be structured and stored in a variety of file formats.

The main goal of research data sharing is to enable other researchers to reuse data. Thus, reusability should always be taken into account when designing systems that create and store research data. We believe that data reuse could be optimized by aligning the 10 aspects of data listed below, Figure 1. This pyramid[3] – loosely modeled on Maslow's hierarchy of human

[1] OMB Circular 110, https://www.whitehouse.gov/omb/fedreg_a110-finalnotice
[2] From 'Defining Research Data' by the University of Oregon Libraries: Documents (text, Word), spreadsheets; Laboratory notebooks, field notebooks, diaries; Questionnaires, transcripts, codebooks; Audiotapes, videotapes Photographs, films; Protein or genetic sequences; Spectra; Test responses; Slides, artifacts, specimens, samples; Collection of digital objects acquired and generated during the process of research; Database contents (video, audio, text, images); Models, algorithms, scripts; Contents of an application (input, output, logfiles for analysis software, simulation software, schemas); Methodologies and workflows; and, Standard operating procedures and protocols.
[3] See figure in '10 aspects of highly effective research data' at https://www.elsevier.com/connect/10-aspects-of-highly-effective-research-data.

needs – can be seen as an extension of the FAIR Data Principles[4] (data should be Findable, Accessible, Interoperable and Reusable) and can function as a roadmap for the development of better data management processes and systems throughout the data lifecycle.

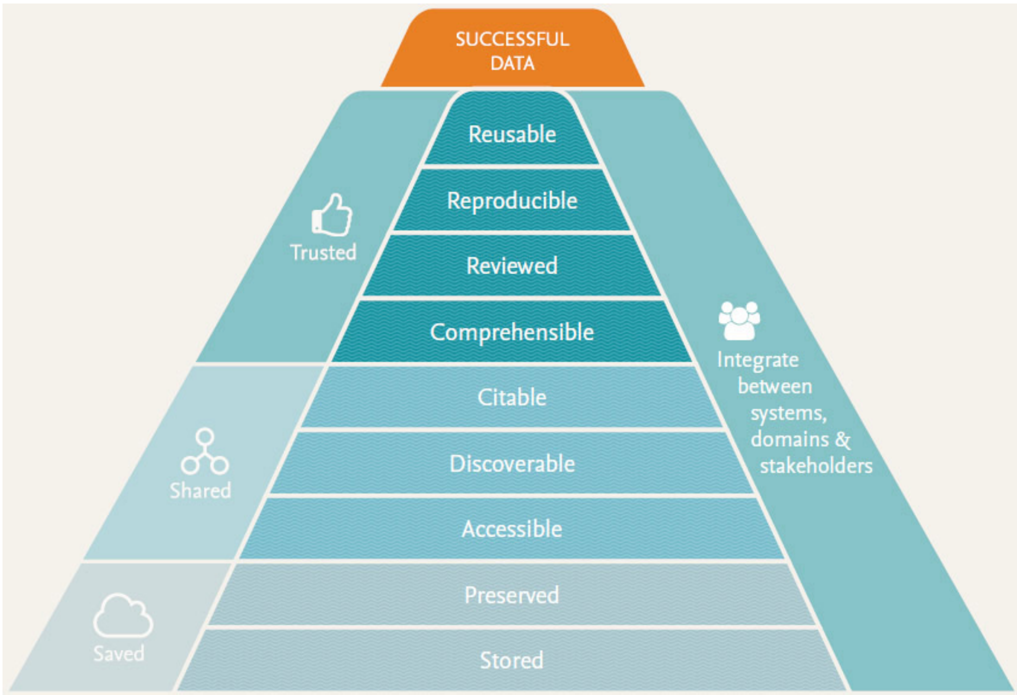

Figure 1: This pyramid can function as a roadmap for the development of better data management processes and systems.

Disciplinarity of Data

While this RFI specifically indicates *biomedical* repositories, it is important to recognize the increasingly interdisciplinary nature of biomedical, life sciences, and health sciences research and the overlaps of research data types from other disciplines.

In a parallel effort, the NSF has been focused on open data and research data through the Open Data Workshop Series[5], the first of which was held in November, 2015. While the workshop's initial focus was on generating discipline-specific responses from the Mathematical and Physical Sciences research communities to the federal policy requiring open data and the recently-released NSF policy statement on open data, there is considerable alignment with the NIH biomedical domain as it relates to research data: decide how and what to preserve in terms of research data for public consumption; the manner by which research data will be stored and accessed; and, the level of burden implied by conservation that is placed on the individual investigator.

[4] Force11 The FAIR Data Principles, https://www.force11.org/group/fairgroup/fairprinciples
[5] NSF MPS Open Data Workshop Series, https://mpsopendata.crc.nd.edu/index.php

International standards organizations, such as the National Data Service (NDS)[6], Research Data Alliance (RDA),[7] and ICSU-World Data System (WDS)[8], have been leading the charge to develop consensus and standards related to research data across disciplines. Elsevier, along with other publishers and research information providers, and additional research ecosystem stakeholders have been working in close partnership with these organizations, and have been engaged with the NSF initiative, as well as working with NIST[9]. These joint efforts have already begun to make significant strides in defining how to publish, find, and reuse research data. We thus recommend that the NIH also participate in this collaborative approach to:

1. Adopt flexible, broad standards and principles related to research data so that all disciplines have the maximum opportunity to interpret research data metrics and demonstrate research impact according to their field and across domains;

2. Consider how to combine quantitative with qualitative inputs; this to ensure that all disciplines, and all agencies and institutions regardless of their disciplinary focus, can share and interpret outcomes and research impact in a similar way;

3. Highlight the full range of types of research data deposit and reuse relevant to many research disciplines, so researchers have the widest opportunity to demonstrate maximum research impact of their work.

Research Metrics

This response focuses on research data, which constitutes an important part of the comprehensive ecosystem of research recognition. We would like to note the following types of research impact that should be considered across the research workflow (Figure 2, below):

1. Research activity – production of outputs leading to enhanced knowledge and understanding, such as original research in journal publications and books, research data, reports, designs, software, etc.; securing income to support ongoing research activities.

2. Research impact – recognition of the influence of research activity on subsequent research through viewing activity, and the receipt of citations from that subsequent research.

---

[6] NDS, http://www.nationaldataservice.org/
[7] Research Data Alliance, https://rd-alliance.org/
[8] ICSU-WDS, https://www.icsu-wds.org/
[9] Public Access to NIST Research, https://www.nist.gov/open

3. Scholarly impact – the wider recognition of research, beyond citing previous work, within the scholarly community, such as the receipt of prizes, requests to edit a journal and to peer review funding applications, and so on.

4. Economic impact – the production of commercializable outputs such as registered and granted patents and spin-out companies, and income generated from these outputs.

5. Social impact – the achievement of societally relevant outcomes, the enhancement of well-being to society as a result of research outputs and/or activities.

A well-rounded, inclusive recognition system can be assessed on all of the facets mentioned above, including research data, by the responsible use of research metrics as good approximations (proxies) of the actual level of performance. The research metrics that are selected should be complemented by the occasional use of narrative inputs such as case studies, firstly as a sanity check that the research metrics are indeed reflective of performance, and secondly in cases where research metrics cannot capture the full value of the research output or outcome.

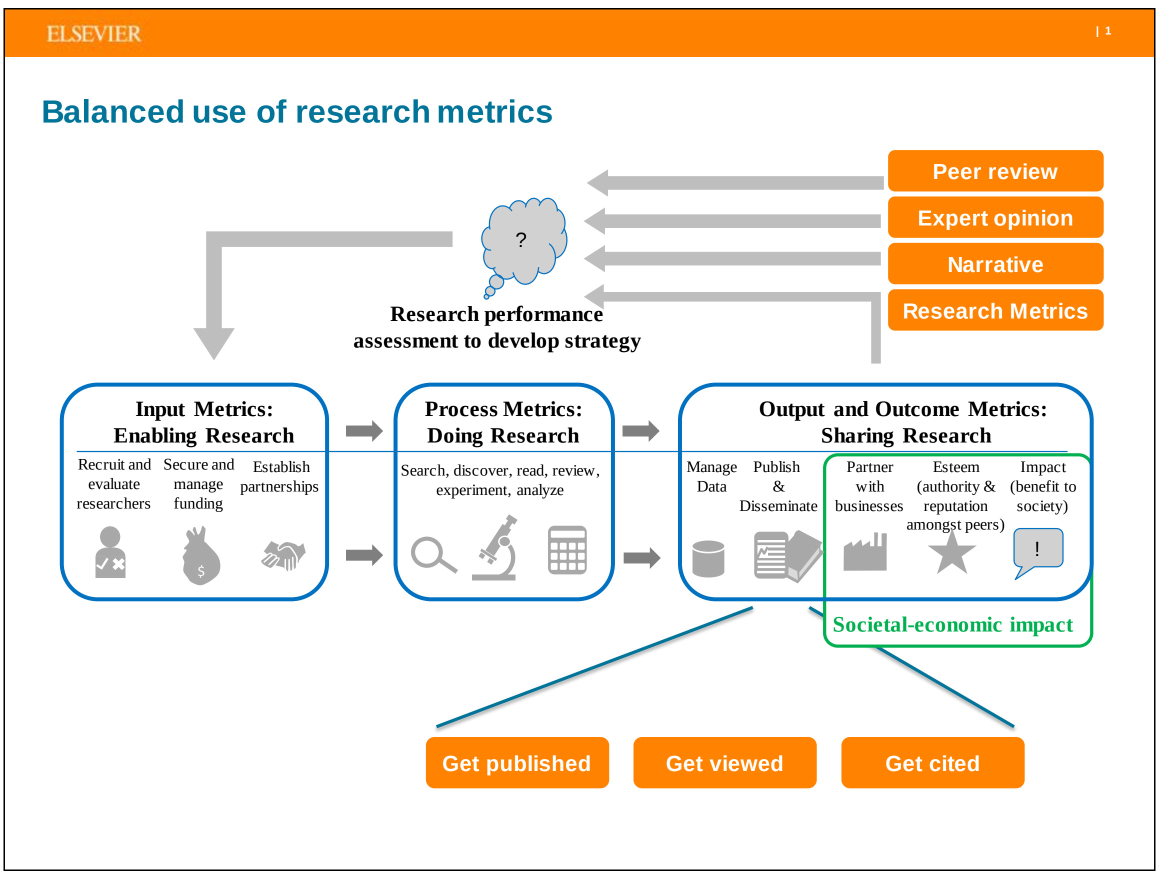

Figure 2: Balanced use of research metrics across the full research workflow.

Golden Rules

Elsevier's work with the research community has led us to recommend two "Golden Rules" for working with research metrics:

1. All decisions and participants benefit from a combination of both quantitative indicators and qualitative (e.g., case studies) input;

2. Quantitative input should always be based on at least two metrics (refer to Table 1 and Table 2 below for examples).

These Golden Rules are a practical reflection of the fact that the highest confidence in decision making is achieved when based on the most complete picture possible, which in turn depends on diverse inputs. Indicators reflect a version of the complete truth that is represented in research data repositories, and as such are an effective proxy for performance. The combination of these indicators can create a good impression of a comprehensive picture, as when a jigsaw has enough pieces in place to gain a good impression of the image, but the indicator jigsaw retains gaps, even when the underlying data sources are comprehensive and a broad set of indicators are used. Consequently, we recommend always complementing quantitative input from indicators with qualitative input from narratives to bring the view into sharper focus, and equally, we recommend that qualitative inputs are always used in combination with indicators.

Basket of Metrics

In close partnership with the research community, we have developed a 'basket of metrics' approach to using research metrics representing all types of research activity across the research workflow (Figure 2); research data metrics are no exception. In the next section, we list research data metrics that would be useful to help measure research impact, but would like to make some general comments about the advantages of an approach that builds on a multiplicity of research metrics here. The advantages of a 'basket of metrics' are:

1. Research excellence, even in one area such as research data, covers a broad range of concepts, and this diversity is best captured by considering a broad range of research metrics.

2. Funders and institutions need flexibility to determine the most appropriate research metrics to demonstrate research impact.

3. The set of research metrics offered can be read out in different ways, which accommodates the expectation by the research community for both simple research

metrics and more sophisticated, but complex, ones. Our research[10] shows that both types are needed and appreciated by users, and both types are important in offering the most complete picture of performance.

a. Simple research metrics such as total counts of activity, and counts normalised by university or faculty size (expressing the indicator as a proportion (%) of total, or by dividing the total count by number of researchers or outputs), are useful for offering transparency and clarity on the underlying data, and for showing the magnitude of activity in absolute terms.

b. More complex research metrics, such as field-normalised algorithms, take into account different behavior between fields and so enable the fair comparison of relative performance in physics with that in biology, for instance.

4. Our work with the community has led us to recommend Two Golden Rules of using research metrics. We discussed the first, always using quantitative measurements together with qualitative inputs, in question 4. The second Golden Rule is to always use at least two quantitative indicators as input into any decision. We recommend that any instance of research impact is demonstrated by using at least two research metrics, because:

a. Every single indicator has its weaknesses as well as its strengths, and these weaknesses can be complemented, or balanced, by the strengths of other indicators.

b. It reduces the likelihood of game playing. There is not, and will never be, one single research metrics that encompasses all aspects of excellent performance. If we try to reduce excellent performance to any single research metric, we will almost certainly drive unbalanced, undesirable behaviour; the researchers could work out how to optimise their performance according to that one research metric. It is much more difficult to see how researchers could adjust their behaviour when the outcomes of that behaviour are measured by using two, or three, or five different research metrics, except by doing genuinely better research across a range of outcomes – which is a result that the NIH is aiming to encourage.

[10] Extensive user research is represented in L. Colledge and C. James, 2015, A “basket of metrics”—the best support for understanding journal merit, European Science Editing 41(3), p61-65; http://europeanscienceediting.eu/articles/a-basket-of-metrics-the-best-support-for-understanding-journal-merit/

Metrics for Research Data

According to the Digital Curation Centre (DCC)[11], "a key measure of the worth of research is the impact it has or, to put it another way, the difference it is making both within the academic community and beyond." It is therefore in the interests of researchers, institutions, and funders to track the impact of research, starting with the impact of research outputs. Historically, research output used to evaluate impact was primarily peer-reviewed research articles. In recent years, other forms of research output are being recognized. The NIH now identifies research data as a legitimate type of 'research product' that can be listed in the "Contributions to Science" section of biosketches submitted as part of a grant application, carrying equal weight with publications.

Elsevier, through Scopus, is leading the way in displaying and collecting journal, article, and author level metrics around scientific literature[12], and intends to do the same for research data (see more below in the "Citation in Practice – The Scopus Model" section). Elsevier's Metrics team, with input from members of the NIH Big Data to Knowledge (BD2K) team, has developed an initial set of quantitative research data metrics (Table 1 and Table 2). All of the research data metrics presented in both tables can be calculated at multiple levels of aggregation (e.g., institution or discipline).

Table 1: Types of Research Data Metrics

Category	Research Data Metric	Description
Collaboration	Collaboration	Proportion of research data outputs with international, or national, or institutional, or no co-authors
Posting	Research Data Outputs	Total count of research data outputs
Get Viewed	Search Count	Total count of times research data outputs have been returned in a search
Get Viewed	Views Count	Total count of views
Get Viewed	Views Percentile measurement	For an individual piece of research data, this would be its percentile according to views received, compared to similar research data outputs For an aggregate entity like an institution, this will be proportion of research data outputs that fall into the top 1%, 5%, 10% or 25% of the world of research data outputs
Get Cited	Citation Count	Total count of citations

[11] Why measure the impact of research data?, http://www.dcc.ac.uk/resources/how-guides/track-data-impact-metrics#why-measure-the-impact-of-research-data
[12] Scopus metrics, https://www.elsevier.com/solutions/scopus/features/metrics

Get Cited	Cited Research Data Outputs	Proportion of Research Data Outputs that have been cited at least once
Get Cited	Citations Percentile measurement	As Views Percentile Measurement
Economic Impact	Academic-Corporate Collaboration	Proportion of research data outputs with both academic and corporate co-authors
Scholarly Impact	Scholarly Activity	This is the total of Mendeley deposits, CiteULike deposits, and similar kind of activity. You can then slice and dice by each individually
Scholarly Impact	Scholarly Commentary	Total mentions in e.g. F1000. You can then slice and dice by each individually
Social Impact	Social Activity	This is the total of Tweets, Facebook likes, and similar kind of activity. You can then slice and dice by each individually
Social Impact	Mass Media	Total mentions in mass media. There are a few variants of this metric we have worked on for publications and which could be applied

Table 2: Research Data Repository Metrics

Category	Research Data Metric	Description
Data Reuse	Data Linkage	Proportion of papers with research data associated with them
Data Reuse	Data Depositing	Proportion of researchers that deposit research data within a certain time frame

Part 2: Research Data Repositories

Defining Trustworthiness

Elsevier has been actively working in robust and deep partnership with numerous national and international research data organizations developing standards for research data repositories. These organizations have made significant strides in defining the criteria that should be used to develop and certify trusted research data repositories.

The most advanced existing data repository certification schemes are:

- Data Seal of Approval (DSA)[13]
- World Data Scheme (WDS) Certification[14]
- Trusted Repositories Audit & Certification (TRAC)[15]

[13] DSA, http://www.datasealofapproval.org/en/
[14] WDS Certification, https://www.icsu-wds.org/services/certification

- Digital Curation Centre (DCC)'s Nestor Catalogue of Criteria for Trusted Digital Repositories[16]

DSA and WDS, whose schemas both rely on self-assessment, are combining their efforts through the Research Data Alliance (RDA)'s Repository Audit and Certification DSA–WDS Partnership Working Group[17] for "realizing efficiencies, simplifying assessment options, stimulating more certifications, and increasing impact on the community. The output from this WG is envisioned as a possible first step towards developing a common framework for certification and a service of trusted data repositories."

DSA includes 16 guidelines[18] covering data producers, data repositories, and data consumers. DSA already has a process in place for the full range of research data repositories to obtain certification, and it maintains a directory of repositories that have successfully acquired certification. The developing DSA-WDS Common Requirements[19] creates a harmonized set of criteria for certification of repositories at the core level addressing research data repository sustainability issues in the areas of organizational infrastructure, digital object management, technology, financial, and legal, etc. Furthermore, the DSA-WDS joint initiative plans to collaborate on a global framework for repository certification that moves from the core to the extended (NESTOR-Seal[20]), to the formal (ISO 16363[21]) level.

Rather than constructing schemas anew specific to biomedical repositories, the current DSA and WDS guidelines and developing Common Requirements must be applied to biomedical repositories to ensure the greatest potential for discoverability and reuse of research data that results from NIH-funded studies and other biomedical research.

Obtaining Certification

From its inception, Elsevier has incorporated the guidance developed by the aforementioned organizations into the development of our multidisciplinary data repository, **Mendeley Data**[22].

---

[15] TRAC, http://www.dcc.ac.uk/resources/repository-audit-and-assessment/trustworthy-repositories
[16] DCC Nestor Catalogue of Criteria for Trusted Digital Repositories, http://www.dcc.ac.uk/resources/repository-audit-and-assessment/nestor
[17] Repository Audit and Certification DSA–WDS Partnership WG, https://rd-alliance.org/groups/repository-audit-and-certification-dsa%E2%80%93wds-partnership-wg.html
[18] DSA Guidelines, http://www.datasealofapproval.org/en/information/guidelines/
[19] DSA-WDS Common Requirements, https://rd-alliance.org/system/files/DSA%E2%80%93WDS%20Catalogue%20of%20Common%20Requirements%20V2.2.pdf
[20] NESTOR Seal for Trustworthy Digital Archives, http://www.langzeitarchivierung.de/Subsites/nestor/EN/nestor-Siegel/siegel_node.html
[21] ISO 16363 Trusted Digital Repositories Management Systems, http://anab.org/programs/isoiec-17021/ms-accreditation-programs/digital-repositories-iso-16363/
[22] Mendeley Data, https://data.mendeley.com/

A critical and absolute criterion of a trusted repository, but one often overlooked by many data repositories, is a mechanism for *long-term* preservation of digital assets. Elsevier has long been a leader in the area of permanent e-journal preservation and an advocate of publisher and research information provider responsibility for digital archiving. Just as Elsevier has done for content published in our journals, we teamed up with DANS (Data Archiving and Networking Services)[23] to ensure that all research datasets within Mendeley Data will be sent offsite to DANS, where they will ensure that the research data is safely archived.

Elsevier is also in the process of obtaining the Data Seal of Approval for Mendeley Data.

Part 3: Data Discoverability

Data Indexing

Elsevier's DataSearch[24] is a prototype research data search engine developed by Elsevier's Research Data Management team that allows users to search for research data across domains and types, from domain-specific, cross-domain, and institutional data repositories. The tool is an exploration of what a search engine for research data needs to look like (versus a web search engine or a document search engine). DataSearch currently indexes images, tables and supplementary data from content sources[25], considered 'research data components.' DataSearch also indexes a series of domain-specific repositories, as well as non-domain specific ones[26]. We are exploring how we might integrate DataSearch with our other offerings, such as Mendeley Data, Scopus, and Pure, to provide robust research data management solutions across the research workflow. And for both, we are working with BD2K on the inclusion of Mendeley Data and DataSearch into the NIH Data Commons.

DataSearch harvests data through APIs (application program interfaces) from various repositories or, in some cases, through database dump files provided to the project. We then normalize the data to our data model, index the data to make it searchable, and generate previews of data where possible. Users can go directly to the source repository from the preview page.

[23] DANS, https://dans.knaw.nl/en

[24] Elsevier DataSearch, https://datasearch.elsevier.com/

[25] Other than from Elsevier's ScienceDirect, DataSearch only indexes open data from open access repositories

[26] As of June 2016, DataSearch is indexing the following content sources: Tables, figures and supplementary data associated with papers in ScienceDirect, arXiv and PubMed Central; Mendeley Data; NeuroElectro; Dryad; PetDB; ICPSR; Harvard Dataverse; and ThemoML at NIST Thermodynamic Research Center (TRC). We are currently investigating DataSearch being able to index all of the NIH-supported data repositories (see https://www.nlm.nih.gov/NIHbmic/nih_data_sharing_repositories.html for list). We will continue to add more content sources in the future.

Elsevier uses a pilot set of criteria to select repositories to index in DataSearch, including the number of users, the ease of our ability to index the repository data, and relationships we have with data repository managers. We are committed to indexing all 63 NIH-supported repositories[27] in DataSearch; we cannot do them all at once, however, so we will seek input from the NIH on ranking/prioritization.

We are also engaging with data repositories to investigate how we can most effectively combine efforts regarding data discovery options, including having DataSearch power search on the repositories themselves. The DataSearch team is working with the NIH-funded bioCADDIE (biomedical and healthCAre Data Discovery Index Ecosystem)[28] team, which has been developing a data discovery index prototype[29] that indexes data that are stored elsewhere, and Elsevier is exploring how we can better collaborate through shared interfaces and API's.

Data Citation

For data to be discovered and acknowledged it must be widely accessible and cited in a consistent and clear manner in the scientific literature. Elsevier endorses the Joint Declaration of Data Citation Principles[30], which will render research data an integral part of the scholarly record, properly preserved and easily accessible, ensuring that researchers get proper credit for their work. The citation principles focus on Importance, Credit and Attribution, Evidence, Unique Identification, Access, Persistence, Specificity and Verifiability, and Interoperability and Flexibility. A data citation is included in the standard References list of an article, and treated on equal footing with article citations.

In Elsevier's ScienceDirect platform, this means readers will enjoy the same benefits with data as they do with article citations, including one-click deep links to the referenced material and the ability to quickly jump to the point in the article where the work was first cited (see Figure 3 below).

[27] NIH Data Sharing Repositories, https://www.nlm.nih.gov/NIHbmic/nih_data_sharing_repositories.html
[28] bioCADDIE, https://biocaddie.org/about
[29] DataMed, https://datamed.org/
[30] Joint Declaration of Data Citation Principles, https://www.force11.org/group/joint-declaration-data-citation-principles-final

References

Barnett et al., 2013 C.L. Barnett, N.A. Beresford, L.A. Walker, M. Baxter, C. Wells, D. Copplestone
Element and radionuclide concentrations in representative species of the ICRP's reference animals and plants and associated soils from a forest in North-west England
NERC — Environmental Information Data Centre (2013) http://dx.doi.org/10.5285/e40b53d4-6699-4557-bd55-10d196ece9ea

Beresford, 2010 N.A. Beresford
The transfer of radionuclides to wildlife (Editorial)
Radiat Environ Biophys, 49 (2010), pp. 505–508
View Record in Scopus | Full Text via CrossRef | Citing articles (10)

Beresford et al., 2008a N.A. Beresford, M. Balonov, K. Beaugelin-Seiller, J. Brown, D. Copplestone, J.L. Hingston, *et al.*
An international comparison of models and approaches for the estimation of the radiological exposure of non-human biota
Appl Radiat Isot, 66 (2008), pp. 1745–1749
Article | PDF (272 K) | View Record in Scopus | Citing articles (27)

Figure 3: The image shows a reference list from the article "A new approach to predicting environmental transfer of radionuclides to wildlife: A demonstration for freshwater fish and caesium," published in *Science of the Total Environment* 2013.

Citation in Practice – The Scopus Model

Elsevier's Scopus is the largest abstract and citation database of peer-reviewed literature: scientific journals, books/book chapters, and conference proceedings. Delivering a comprehensive overview of the world's research output in the fields of science, technology, medicine, social sciences, and arts and humanities, Scopus features smart tools to track, analyze, and visualize research and its impact. Scopus' vision of research data aligns with the Force11 Joint Declaration of Data Citation Principles[31] which state that research data is as integral to recognizing and assessing the research output of modern researchers as are articles, reviews, books and all other "traditional" forms of research output (refer to Figure 2). Thus, research data must be:

- Discoverable
- Trustworthy
- Included in the author profile
- Creditable

DataSearch and Scopus are taking a complementary approach. Whereas DataSearch indexes a number of data sources and allows researchers to discover, access, and preview relevant data sets in multiple formats, the goal for Scopus is to integrate and curate DataSearch results to ensure that the research data discoverable via Scopus.com is trustworthy, in a manner consistent with the approach we take toward traditional content inclusion by way our independent Content Selection & Advisory Board (CSAB)[32].

---

[31] https://www.force11.org/group/joint-declaration-data-citation-principles-final

[32] Scopus CSAB, https://www.elsevier.com/solutions/scopus/content/scopus-content-selection-and-advisory-board

Presently the Scopus CSAB vets all journals indexed in Scopus to ensure high quality standards. We believe that a similar methodology should be applied to data repositories, to ensure transparent, consistent, high quality content

Integrating a research data search engine such as DataSearch in Scopus as a prototype will require a combination of human and algorithmic curation techniques to ensure that Scopus users can trust and rely on the results. In order to achieve this, we intend to apply rigorous selection criteria to both data repositories and data types (refer to the sections on "Research Data Definition" and "Defining Trustworthiness" above for criteria that we will consider).

After ensuring research data is discoverable, the next step will be for Scopus to integrate research data citations in Scopus Author Profiles, to appropriately link and assign credit to the author. Metrics can be applied to research data citations in Scopus just as they are now for articles (refer to the section above, "Metrics for Research Data").

Scopus is leading the way in displaying and collecting journal, article, and author level metrics around scientific literature[33], and intends to do the same for research data. Several parameters will be developed to attribute metrics to data. Scopus will collect and display these metrics in a way that is clear and imparts meaning and value to each metric. Through these efforts, Elsevier can enhance recognition across the research workflow (Figure 2) through enhancement of data search and credit for research data output.

Part 4: Recognition and Reward

While this RFI doesn't specifically identify the topic recognition and reward of research data to support widespread research data sharing, we think that the issue is inextricably linked to the sustainability of research data repositories.

At the SciDataCon 2016 conference in September, 2016, there was a session entitled, *Getting the incentives right: Removing social, institutional and economic barriers to data sharing*[34]. The session description indicates that while "much work has been done relating to the technical aspects of scientific data sharing…[progress toward research data sharing]…has been particularly hampered by a lack of awareness that the barriers and risks to be addressed are socio-technical concerns, with the non-technical concerns –the social, institutional and economic aspects of data sharing, often overlooked."

[33] Scopus metrics, https://www.elsevier.com/solutions/scopus/features/metrics
[34] http://www.scidatacon.org/2016/sessions/37/

Elsevier has been working with the research data community to compile a body of literature addressing the socio-technical aspects of research data sharing rewards and incentives, as well as relevant references on knowledge sharing incentive systems (Table 3)[35]. We recommend that this literature be comprehensively evaluated with the goal of developing recommendations for effective policies and practices that the NIH (and other funders), research institutions, and faculty promotion & tenure committees can employ to promote research data sharing.

Table 3: References on Rewards and Incentives for Research Data Sharing

1. Anderson MS, Ronning E a., De Vries R, Martinson BC. The perverse effects of competition on scientists' work and relationships. Sci Eng Ethics. 2007;13:437–61.
2. Arzi S, Rabanifard N, Nassajtarshizi S, Omran N. Relationship among Reward System, Knowledge Sharing and Innovation Performance. Interdiscip J Contemp Res Bus. 2013;5(6):115–41.
3. Bartol K. Encouraging Knowledge Sharing: The Role of Organizational Reward Systems. J Leadersh & Organ Stud. 2002;9(1):64–76.
4. Birney E, Hudson TJ, Green ED, Gunter C, Eddy S, Rogers J, et al. Prepublication data sharing. Nature [Internet]. 2009;461(7261):168–70. Available from: http://dx.doi.org/10.1038/461168a
5. Borgman CL. The Conundrum of Sharing Research Data. SSRN Electron J [Internet]. 2011;63(6):1–40. Available from: http://www.ssrn.com/abstract=1869155
6. Boudreau KJ, Lakhani KR. "Open" disclosure of innovations, incentives and follow-on reuse: Theory on processes of cumulative innovation and a field experiment in computational biology. Res Policy. 2015;44(1):4–19.
7. Bourne PE, Lorsch JR, Green ED. OUTLOOK BIG DATA IN BIOMEDICINE Sustaining the big-data ecosystem. 2015;
8. Carrara W, Fischer S, Steenbergen E van. Open Data Maturity in Europe 2015: Insights into the European state of play. European Data Portal Open. 2015.
9. Chia-Shen C, Shih-Feng C, Chih-Hsing L. Understanding Knowledge-Sharing Motivation, Incentive Mechanisms, and Satisfaction in Virtual Communities. Soc Behav Personal An Int J [Internet]. 2012;40(4):639–47. Available from: http://proxy.indianatech.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=s3h&AN=75245509&site=ehost-live&scope=site
10. Costello MJ. Motivating Online Publication of Data. Bioscience [Internet]. 2009;59(5):418–27. Available from: http://www.jstor.org/stable/10.1525/bio.2009.59.5.9
11. Cress U, Barquero B, Schwan S, Hesse FW. Improving quality and quantity of contributions: Two models for promoting knowledge exchange with shared databases. Comput Educ [Internet]. 2007 [cited 2016 Sep 15];49(2):423–40. Available from: http://www.sciencedirect.com/science/article/pii/S0360131505001375
12. Denis J, Goëta S. Exploration, Extraction and "Rawification". The Shaping of Transparency in the Back Rooms of Open Data. Soc Sci Res Netw [Internet]. 2014; Available from: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2403069&download=yes
13. Edwards PN, Mayernik MS, Batcheller AL, Bowker GC, Borgman CL. Science friction: Data, metadata, and collaboration. Soc Stud Sci [Internet]. 2011 Aug 15 [cited 2012 Mar 22];41(5):667–90. Available from: http://sss.sagepub.com/cgi/content/abstract/0306312711413314v1
14. Ember C, Hanisch R, Alter G, Berman H, Hedstrom M, Vardiagn M. Sustaining Domain Repositories for Digital Data: A White Paper. 2013;(February):1–16.
15. Enke N, Thessen A, Bach K, Bendix J, Seeger B, Gemeinholzer B. The user's view on biodiversity data sharing - Investigating facts of acceptance and requirements to realize a sustainable use of research data -. Ecol Inform. 2012;11:25–33.
16. Fecher B, Friesike S, Hebing M. What Drives Academic Data Sharing? SSRN Electron J [Internet]. Berlin, Germany; 2014;10(2). Available from: http://papers.ssrn.com/abstract=2439645
17. Fecher B, Friesike S, Hebing M, Linek S, Sauermann A. A Reputation Economy: Results from an Empirical Survey on Academic Data Sharing [Internet]. Berlin, Germany; 2015. Report No.: 1454. Available from: http://d.repec.org/n?u=RePEc:diw:diwwpp:dp1454&r=sog
18. Fitch P, Craglia M, Pollock R, Cox S, Fowler D. Getting the incentives right: removing social, institutional and economic barriers to data sharing [Internet]. International Data Week Conference Session. Denver, CO, USA; 2016. Available

[35] Holly Falk-Krzesinski, PhD, at Elsevier can be contacted directly to be added to the growing reference group, h.falk-krzesinski@elsevier.com

from: http://www.scidatacon.org/2016/sessions/37/
19. Friesike S, Fecher B, Hebing M, Linek S. Reputation Instead of Obligation : Why We Need to Forge New Policies to Motivate Academic Data Sharing [Internet]. Blog Post. Alexander von Humboldt Institute for Internet and Society; 2015. p. 5–8. Available from: http://www.hiig.de/en/23202/
20. Gardner D, Toga AW, Ascoli G a, Beatty JT, Brinkley JF, Dale AM, et al. Towards effective and rewarding data sharing. Neuroinformatics. 2003;1(3):289–95.
21. Gaulé P, Maystre N. Getting cited: Does open access help? Res Policy. 2011;40(10):1332–8.
22. Goëta S. Instaurer des données, instaurer des publics- Une enquête sociologique dans les coulisses de l'open data: Instantiate data, instantiate publics- a sociological inquiry in the backrooms of open data. [Paris, France, France]: Télécom Paris Tech; 2016.
23. Gorgolewski KJ, Margulies DS, Milham MP. Making data sharing count: A publication-based solution. Front Neurosci. 2013;(7 FEB).
24. Hung SY, Durcikova A, Lai HM, Lin WM. The influence of intrinsic and extrinsic motivation on individuals knowledge sharing behavior. Int J Hum Comput Stud. 2011;69(6):415–27.
25. Ingram C. How and why you should manage your research data: a guide for researchers [Internet]. JISC; 2016. Available from: https://www.jisc.ac.uk/guides/how-and-why-you-should-manage-your-research-data?mkt_tok=3RkMMJWWfF9wsRonuqjMZKXonjHpfsX56%2B4pW6S%2BlMI%2F0ER3fOvrPUfGjI4ATMRmI%2BSLDwEYGJlv6SgFTrLHMa1izLgNUhA%3D
26. Jahani. Is Reward System and Leadership Important in Knowledge Sharing Among Academics? American Journal of Economics and Business Administration. 2011. p. 87–94.
27. Kim Y, Adler M. Social scientists' data sharing behaviors: Investigating the roles of individual motivations, institutional pressures, and data repositories. Int J Inf Manage. 2015;35(4):408–18.
28. Koers H. How do we make it easy and rewarding for researchers to share their data? – a publisher's perspective. J Clin Epidemiol [Internet]. 2015 Jul [cited 2015 Jul 14]; Available from: http://www.sciencedirect.com/science/article/pii/S089543561500325X
29. Li Y-M, Jhang-Li J-H. Knowledge sharing in communities of practice: A game theoretic analysis. Eur J Oper Res [Internet]. 2010;207(2):1052–64. Available from: http://www.sciencedirect.com/science/article/pii/S0377221710003899
30. Lin S-W, Lo LY-S. Mechanisms to motivate knowledge sharing: integrating the reward systems and social network perspectives. J Knowl Manag [Internet]. 2015;19(2):212–35. Available from: http://www.emeraldinsight.com.ezp.lib.unimelb.edu.au/doi/full/10.1108/JKM-05-2014-0209
31. Longo DL, Drazen JM. Data Sharing. N Engl J Med [Internet]. 2016;374(3):276–7. Available from: http://dx.doi.org/10.1056/NEJMe1516564\nhttp://www.nejm.org/doi/full/10.1056/NEJMe1516564
32. Mayernik MS, Callaghan S, Leigh R, Tedds J, Worley S. Peer Review of Datasets: When, Why, How. Bull Am Meteorol Soc. 2014;1–32.
33. Mueller-Langer F, Andreoli-Versbach P. Open Access to Research Data: Strategic Delay and the Ambiguous Welfare Effects of Mandatory Data Disclosure. Berlin, Germany; 2014. Report No.: 239.
34. Muller R, Spiliopoulou M, Lenz H. The Influence of Incentives and Culture on Knowledge Sharing. System Sciences, 2005 HICSS '05 Proceedings of the 38th Annual Hawaii International Conference on [Internet]. 2005. p. 247b--247b. Available from: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1385745
35. Nelson B. Data sharing: Empty archives. Nature [Internet]. 2009;461:106–63. Available from: http://www.nature.com/news/2009/090909/full/461160a.html
36. Niu JJ. Reward and Punishment Mechanism for Research Data Sharing. IASSIST Q. 2006 May;(Winter).
37. Pham-Kanter G, Zinner DE, Campbell EG. Codifying collegiality: Recent developments in data sharing policy in the life sciences. PLoS One [Internet]. 2014;9(9). Available from: http://journals.plos.org/plosone/article/asset?id=10.1371/journal.pone.0108451.PDF
38. Pisani E, AbouZahr C. Sharing health data: good intentions are not enough. Bull World Health Organ [Internet]. 2010;88(6):462–6. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2878150&tool=pmcentrez&rendertype=abstract
39. Piwowar HA, Chapman WW. A review of journal policies for sharing research data. Open Scholarship: Authority, Community, and Sustainability in the Age of Web 20 - Proceedings of the 12th International Conference on Electronic Publishing, ELPUB 2008 [Internet]. 2008. p. 1–14. Available from: http://elpub.scix.net/cgi-bin/works/Show?001_elpub2008
40. Poline J-B, Breeze JL, Ghosh S, Gorgolewski K, Halchenko YO, Hanke M, et al. Data sharing in neuroimaging research. Front Neuroinform [Internet]. 2012;6:9. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3319918&tool=pmcentrez&rendertype=abstract
41. Ranganathan K, Ripeanu M, Sarin a., Foster I. Incentive mechanisms for large collaborative resource sharing. IEEE Int Symp Clust Comput Grid, 2004 CCGrid 2004 [Internet]. 2004;1–8. Available from:

http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1336542

42. Roche DG, Kruuk LEB, Lanfear R, Binning SA. Public Data Archiving in Ecology and Evolution: How Well Are We Doing? PLOS Biol [Internet]. 2015;13(11):e1002295. Available from: http://dx.plos.org/10.1371/journal.pbio.1002295
43. Rolland B. Data Sharing and Reuse: Expanding Our Concept of Collaboration [Internet]. Team Science Toolkit Blog. 2016 [cited 2016 Jan 1]. Available from: https://www.teamsciencetoolkit.cancer.gov/Public/ExpertBlog.aspx?tid=4
44. Rood RB, Edwards PN. Climate Informatics: Human Experts and the End-to-end System. Earthzine [Internet]. 2014;1–14. Available from: http://earthzine.org/2014/05/22/climate-informatics-human-experts-and-the-end-to-end-system/
45. Šajeva S. Encouraging Knowledge Sharing among Employees: How Reward Matters. Procedia - Soc Behav Sci [Internet]. 2014;156(April):130–4. Available from: http://www.sciencedirect.com/science/article/pii/S1877042814059540
46. Sarathy R, Muralidhar K. Secure and useful data sharing. Decis Support Syst. 2006;42(1):204–20.
47. Sayogo DS, Pardo TA. Exploring the determinants of scientific data sharing: Understanding the motivation to publish research data. Gov Inf Q. 2013;30(SUPPL. 1).
48. Schofield PN, Bubela T, Weaver T, Portilla L, Brown SD, Hancock JM, et al. Post-publication sharing of data and tools. Nature [Internet]. 2009;461(7261):171–3. Available from: http://dx.doi.org/10.1038/461171a\nhttp://www.nature.com/doifinder/10.1038/461171a
49. Seonghee K, Boryung J. An analysis of faculty perceptions: Attitudes toward knowledge sharing and collaboration in an academic institution. Libr Inf Sci Res. 2008;30(4):282–90.
50. Shuman LJ. Data Sharing in Engineering Education. Adv Eng Educ [Internet]. 2016;5(2). Available from: http://advances.asee.org/summer-2016-volume-5-issue-2/
51. Stanley B, Stanley M. Data sharing: The primary researcher's perspective. Law Hum Behav. 1988;12(2):173–80.
52. Sterling TD, Weinkam JJ. Sharing scientific data. Commun ACM [Internet]. 1990;33(8):112–9. Available from: http://portal.acm.org/citation.cfm?id=79182\nhttp://delivery.acm.org.ezp2.bath.ac.uk/10.1145/80000/79182/p112-sterling.pdf?key1=79182&key2=8630174821&coll=GUIDE&dl=GUIDE&CFID=104901268&CFTOKEN=61271004
53. Tenopir C, Allard S, Douglass K, Aydinoglu AU, Wu L, Read E, et al. Data Sharing by Scientists: Practices and Perceptions. Neylon C, editor. PLoS One [Internet]. 2011 Jun 29 [cited 2011 Jun 30];6(6):e21101. Available from: http://dx.plos.org/10.1371/journal.pone.0021101
54. Thorisson G a. Accreditation and attribution in data sharing. Nat Biotechnol [Internet]. 2009;27(11):984–5. Available from: http://dx.doi.org/10.1038/nbt1109-984b
55. Toronto International Data Release Workshop Authors. Prepublication data sharing. Nature. 2009;461(September):168–70.
56. Van Noorden R. Data-sharing: Everything on display. Nature. 2013;500:243–5.
57. Wallis JC, Rolando E, Borgman CL. If We Share Data, Will Anyone Use Them? Data Sharing and Reuse in the Long Tail of Science and Technology. PLoS One [Internet]. 2013;8(7):e67332. Available from: http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0067332
58. Weber NM, Baker KS, Thomer AK, Chao TC, Palmer CL. Value and context in data use: Domain analysis revisited. Proc Assoc Inf Sci Technol [Internet]. 2012;49(1):1–10. Available from: http://onlinelibrary.wiley.com/doi/10.1002/meet.14504901168/abstract
59. Yang HL, Wu TCT. Knowledge sharing in an organization. Technol Forecast Soc Change. 2008;75(8):1128–56.
60. Zimmerman A. Data Sharing and Secondary Use of Scientific Data: Experiences of Ecologists [Internet] [phdthesis]. 2003. Available from: http://deepblue.lib.umich.edu/dspace/handle/2027.42/39373
61. Zinner D, Pham-Kanter G, Campbell E. The Changing Nature of Scientific Sharing and Withholding in Academic Life Sciences Research: Trends From National Surveys in 2000 and 2013. Acad Med [Internet]. 2016;91(3):433–40. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26675188