Approaching Open Science Across the Researcher Workflow

NASEM Meeting
Toward an Open Science Enterprise: Focus on Stakeholders

September 18, 2017

Holly J. Falk-Krzesinski, VP Research Intelligence
Global Strategic Networks ♦ Elsevier
h.falk-krzesinski@Elsevier.com
Open Science
Open Science Defined

Movement, with the overarching goal of *enhanced research performance*, that aims to make science more:

- Accessible
- Collaborative
- Transparent
- Effective
- Efficient

Through:
- Encouraging a culture of openness and sharing
- Leveraging and developing new technologies
- Developing and adapting reward and metric systems
Publishers are a Key Stakeholder in Open Science

Figure 1 Drivers of open science (Questionnaire responses to ‘What are the key drivers of ‘Science 2.0’?’)

- Availability of digital technologies and their increased capacities: 76% agree, 22% partially agree, 2% disagree.
- Researchers looking for new ways of disseminating their output: 47% agree, 43% partially agree, 2% disagree.
- Researchers looking for new ways of collaboration: 43% agree, 43% partially agree, 3% disagree.
- Increase of the global scientific population: 39% agree, 46% partially agree, 4% disagree.
- Growing criticism of current peer-review system: 34% agree, 42% partially agree, 6% disagree.
- Public demand for better and more effective science: 36% agree, 42% partially agree, 2% disagree.
- Public funding supporting ‘Science 2.0’: 32% agree, 41% partially agree, 6% disagree.
- Growing public scrutiny of science and research: 28% agree, 44% partially agree, 3% disagree.
- Public demand for faster solutions to Societal Challenges: 26% agree, 45% partially agree, 3% disagree.
- Scientific publishers engaging in ‘Science 2.0’: 22% agree, 40% partially agree, 6% disagree.

Citizens acting as scientists:
- 11% totally agree, 33% partially agree, 6% don’t know, 34% partially disagree, 16% totally disagree.

Sample size: 492, missing: 8 to 12.


---

Forum on Open Science

The Forum on Open Science will facilitate discussions on how the full opportunities of opening science might be realized. It will provide a venue for candid, ongoing discussions to identify barriers to open science and determine key questions for industry, academic, and government efforts related to opening science.

The Forum on Open Science will evolve over time, but will aim to serve as a conduit for conversations on the development of open science that will include the following topics:

- Barriers to open science and potential partnerships and initiatives that could help overcome them;
- Potential commercial opportunities in open science;
- Infrastructure systems that will facilitate opening science and models for supporting them;
- Pilot studies to advance broad access to scientific results with a particular focus on how to make scientific data available and useful;
- Impact on commercial, non-profit enterprises, and institutes of higher education of various models of open science; and
- Research needed to fill data gaps on implementation of various open science models.

Staff Director: Meredith Lane

http://sites.nationalacademies.org/DBASSE/CurrentProjects/DBASSE_087866
Open Science Across the Researcher Workflow

“Enabling Research”

- Develop Strategy
- Recruit/evaluate researchers
- Secure Funding
- Establish partnerships
- Manage facilities

“Doing Research”

- Search, discover, read, review
- Collaborate & network
- Experiment
- Synthesize/Analyze

“Sharing Research”

- Manage Data
- Publish and disseminate
- Commercialize
- Promote

Have impact

Open Science opportunity type:
- More open
- More collaborative
- More transparent

- Do ‘office admin’ tasks
- Identify funding opportunities
- Identify potential partners
- Find talent to join team
- Prepare proposals
- Benchmark own performance
- Track proposals
- Manage performance of team
- Report grant progress and compliance

- Stay on top of the field
- Keep track of colleagues/competitors
- Plan experiments, procure materials and resources
- Prepare data for use on different tools
- Analyse results
- Select and format data outputs for presentation
- Promote within institution

- Get in-depth knowledge on a topic
- Attend conferences, build relationships
- Run experiments
- Validate data against prior research
- Prepare data for use on different tools
- Prepare research for publication
- Register a patent
- Communicate results to media

- Search online literature
- Share others’ articles, data
- Share own work in progress
- Public collect data/analyse
- Validate data against prior research
- Present findings at conferences
- Identify commercial partners
- Negotiate partnerships

- Manage articles
- Annotate articles
- Read articles
- Develop hypotheses
- Review manuscripts
- Peer review other’s work
- Monitor progress of collaborative tasks
- Capture, log, and store data
- Find data-sets online and re-use for own research
- Format manuscript for submission
- Prepare teaching materials
- Teach, give lectures, tutoring

- Keep track of when carried out experiments
- Find right methods
- Publish data
- Publish methods
- Find data-sets online and re-use for own research
- Format manuscript for submission
- Prepare teaching materials
- Teach, give lectures, tutoring

- Preserve data
- Describe annotate data
- Publish data
- Present findings at conferences
- Identify appropriate journal
- Format manuscript for submission
- Prepare teaching materials
- Teach, give lectures, tutoring

- Preserve data
- Describe annotate data
- Publish data
- Present findings at conferences
- Identify appropriate journal
- Format manuscript for submission
- Prepare teaching materials
- Teach, give lectures, tutoring

- Preserve data
- Describe annotate data
- Publish data
- Present findings at conferences
- Identify appropriate journal
- Format manuscript for submission
- Prepare teaching materials
- Teach, give lectures, tutoring
Key Elements of Open Science

Developing technology, tools and services that address the challenges of open science

- **Open Access**: Improving access and sharing of research publications
- **Research Data**: Improving access to and use of research data
- **Research Integrity**: Improve reproducibility and transparency of research
- **Science & Society**: Encouraging citizen involvement & translating science for the public
- **Metrics**: Developing metrics which show the full impact of research
Open Access
Open Access By the Numbers*

- 1.2M manuscripts submitted
- 1.8M authors listed on submitted manuscripts
- 800K reviewers (3M potential reviewers)
- 80K journal editors (20K level 1)
- 420K articles published
- 16% article share
- Published 163 of 164 sci & econ Nobel winners since 2000
- 14M scholarly works in ScienceDirect, plus 30K e-books
- 14M ScienceDirect users/mo
- 900M articles downloaded from ScienceDirect

• 24K articles published Gold Open Access (2nd largest Gold OA publisher)
• All 2.5K+ proprietary journals offer a Green Open Access, all support CHORUS
• 3K+ accepted manuscripts available for US partner funding agencies through CHORUS
• 1.55 FWCI*** for our hybrid OA portfolio, highest in the world
• 1.08 FWCI*** for our fully Gold OA journals
• 28K new authors sharing preprints on SSRN, 78K new preprints posted
• 330K authors total sharing on SSRN, 592K preprints total

*2016 Elsevier data unless otherwise noted; **Period 2012-2016; ***2013 data
Open Access at Elsevier

**Gold Open Access**

- Final version of an article is immediately accessible to everyone
- Cost of publishing is recovered upfront, typically as an Article Publishing Charge (APC)

- 2nd Largest gold open access publisher
- Publish over 200 fully open access journals
- 1700+ hybrid journals
- Publish over 20,000 open access articles each year
- Choice of either a commercial (CC BY) or non-commercial (CC-BY-NC-ND) user license
- Article publishing charges (APCs) range from $500- $5000 (US Dollars)

**Green Open Access**

- A version of a subscription article is made available open access, usually after an embargo period
- Normally the accepted manuscript version of an article, can be the final version

- Largest publisher enabler of green open access
- All 2500+ journals provide a green open access option
- Free API program to fuel repositories
- Participate in CHORUS and support pilots with institutions and international funders
- Share link service provides 50 days free access to recently published research
- Open archives in 108 journals, including all Cell Press titles after 12 months
Institutional Repository Services

ScienceDirect

Public Access Services:
• Automatic ingest of embargoes for AMs
• Article versions in line with sharing & hosting policy
• New Pilot Service: Embedded Accepted Manuscripts for Institutional Repositories
• Service: includes a downloadable AM for users after embargo

Scopus

Enhances Your Reputation and Workflows:
• Maximize visibility of research output through automatic API ingest

ScienceDirect

Optimize User Experience/Guide to the Best Article Version:
• Gold OA and Accepted Manuscripts for all users
• Subscription articles for subscribers

Capture Usage and Cited-by Counts for Your Authors

www.elsevier.com/solutions/sciencedirect/support/institutional-repository
“I think this project with Elsevier, and the follow on project with CHORUS, are important and worthwhile efforts to **identify and enhance access to journal articles** by UF authors.

I believe that **automated solutions** for identification and access to articles by UF authors from publishers reduces the burden of gathering this information on our academic faculty and on the library faculty and staff.”

Judith C. Russell, Dean of University Libraries, University of Florida (Aug 2017)
Elsevier Partners with CHORUS

Author submits their manuscript to the journal of their choice

Author identifies their source of funding

Article is accepted after peer review

Article is assigned with a unique ID and funding body ID

Publisher sends this information to Crossref as part of the metadata of the article

ScienceDirect

The article is published on the publisher's platform

Publisher ensures access to the best available version meaning...

Readers of an open access article or researchers affiliated with a subscribing library, will see the final version of the article

Data is fed through to CHORUS dashboards where funders can monitor and track real-time compliance

Article is permanently preserved via CLOCKSS, PORTICO and other 3rd party dark archives

The article can also be text mined via Crossref's TEXT and DATA MINING SERVICE

Article is also indexed and can be found via the CHORUS portal or funder's repository

Article is indexed and can be found by any existing search engine

Other readers will see the full text of the accepted manuscript after an embargo expires

https://www.elsevier.com/about/open-science/open-access/chorus
Research Data
Open Research Data

What is Research Data?

- Refers to the result of observations or experimentation that validate research findings, data that often underlies, but exists outside of research articles.
- Can include but are not limited to: raw data, processed data, software, algorithms, protocols, methods, materials, and which are not already published as part of a journal article.

Why make it open?

Datasets that are freely shared can be valuable for the whole research community and beyond.

- Facilitates the re-use of data for new studies.
- Makes research more transparent and reproducible.
- Makes research more effective and efficient.

https://www.elsevier.com/about/open-science/research-data/open-data-report
Research Data Management Across the Research Data Life Cycle

Source: JISC: How and why you should manage your research data: a guide for researchers, Caroline Ingram, Published: 7 January 2016
Trusted Data Repository

Part of an award to the NIH Data Commons “Pilot Phase”

Store up to 5 GB of data in many formats

Describe how experiment can be reproduced

Create DOI for Citation

Link back to protocols

Keep track of versions of dataset

https://data.mendeley.com/
Research Data Guidelines

• Integrating TOP Guidelines directly into journal workflow making it easy for authors to share the data supporting their articles

• New journal data guidelines that align with the TOP Data Standards, implemented across >1,800 journals

• Integrated into the author submission system, with updated Guides for Authors in the journals
## Research Data Metrics

<table>
<thead>
<tr>
<th>Goal:</th>
<th>Metric:</th>
</tr>
</thead>
<tbody>
<tr>
<td>More data is <strong>saved</strong>:</td>
<td></td>
</tr>
<tr>
<td>1. <strong>Stored</strong>, i.e. safely available in long-term repository)</td>
<td>Nr of datasets stored in long-term storage</td>
</tr>
<tr>
<td>2. <strong>Published</strong>, i.e. long-term preserved, accessible via web, have a GUID, citeable, with proper metadata</td>
<td>Nr of datasets published, in some form</td>
</tr>
<tr>
<td>3. <strong>Linked</strong>, to articles or other datasets</td>
<td>Nr of datasets linked to articles</td>
</tr>
<tr>
<td>4. <strong>Validated</strong>, by a reviewer/curated</td>
<td>Nr of datasets in curated databases/peer reviewed in data articles</td>
</tr>
<tr>
<td>More data is <strong>seen and used</strong>:</td>
<td></td>
</tr>
<tr>
<td>5. <strong>Discovered</strong></td>
<td>Nr of datasets viewed in databases/websites/search engines</td>
</tr>
<tr>
<td>6. <strong>Identified</strong></td>
<td>DOI is resolved</td>
</tr>
<tr>
<td>7. <strong>Mentioned</strong></td>
<td>Social media and news mentions</td>
</tr>
<tr>
<td>8. <strong>Cited</strong></td>
<td>Nr of datasets cited in articles</td>
</tr>
<tr>
<td>9. <strong>Downloaded</strong></td>
<td>Downloaded from repositories</td>
</tr>
<tr>
<td>10. <strong>Reused</strong></td>
<td>Mention of usage in article or other dataset</td>
</tr>
</tbody>
</table>

[https://rdmi.uchicago.edu/papers/08212017144742_deWaard082117.pdf](https://rdmi.uchicago.edu/papers/08212017144742_deWaard082117.pdf)
Additional Resources


References that offer insight and perspective on Open Access and Academic Freedom from faculty members, researchers, editors and authors.

This group is dedicated to references and resources on *research and academic data sharing* to inform motivations and rewards for sharing, maximizing the return on investment for research, increasing transparency, accelerating data-driven innovation, and promoting reproducibility in research.
Additional Resources


Research Data Sharing Group. https://www.mendeley.com/community/research-data-sharing/. This public group on Mendeley is a crowd-sourced, community resource containing an abundant library of references on research and academic data sharing to inform on issues of motivations and rewards for sharing, maximizing the return on investment for research, increasing transparency, accelerating data-driven innovation, and promoting reproducibility in research.

Open Access: Faculty and Researcher Perspectives Group, https://www.mendeley.com/community/open-access-faculty-researcher-perspectives/. This public group on Mendeley contains references that offer insight and perspective on open access and academic freedom from faculty members, researchers, editors and authors.

Report and recommendations from the Scholarly Publishing Roundtable, https://www.aau.edu/sites/default/files/AAU%20Files/Key%20Issues/Intellectual%20Property/Scholarly%20Publishing%20Roundtable%20Report%20and%20Recommendations%20-%201-12-10.pdf. After recognizing the progress that has already been made in expanding access to scholarly literature, the Roundtable began its work by identifying a set of principles, shared across the full range of member perspectives, which should continue to inhere in scholarly publishing as it evolves.

The PEER Project, http://www.peerproject.eu/fileadmin/media/reports/20120618_PEER_Final_public_report_D9-13.pdf. PEER (Publishing and the Ecology of European Research), supported by the EC eContentplus programme, investigated the effects of the large-scale, systematic depositing of authors’ final peer-reviewed manuscripts (so called Green Open Access or stage-two research output) on reader access, author visibility, and journal viability, as well as on the broader ecology of European research. The project was a collaboration between publishers, repositories and researchers and lasted from 2008 to 2012.


The FAIR Data Principles, https://www.force11.org/group/fairgroup/fairprinciples One of the grand challenges of data-intensive science is to facilitate knowledge discovery by assisting humans and machines in their discovery of, access to, integration and analysis of, task-appropriate scientific data and their associated algorithms and workflows. Here, we describe FAIR - a set of guiding principles to make data Findable, Accessible, Interoperable, and Re-usable.

Open Data: The Researcher Perspective, https://www elsevier.com/__data/assets/pdf_file/0004/281920/Open-data-report.pdf. This report is a result of a year-long, co-conducted study between Elsevier and the Centre for Science and Technology Studies (CWTS), part of Leiden University, the Netherlands. The study is based on a complementary methods approach consisting of a quantitative analysis of bibliometric and publication data, a global survey of 1,200 researchers and three case studies including in-depth interviews with key individuals involved in data collection, analysis and deposition in the fields of soil science, human genetics and digital humanities.
Thank you

For more information, please visit: https://www.elsevier.com/about/open-science/research-data/research-data-management and https://www.elsevier.com/about/open-science/research-data/research-data-management/rdm-videos