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:

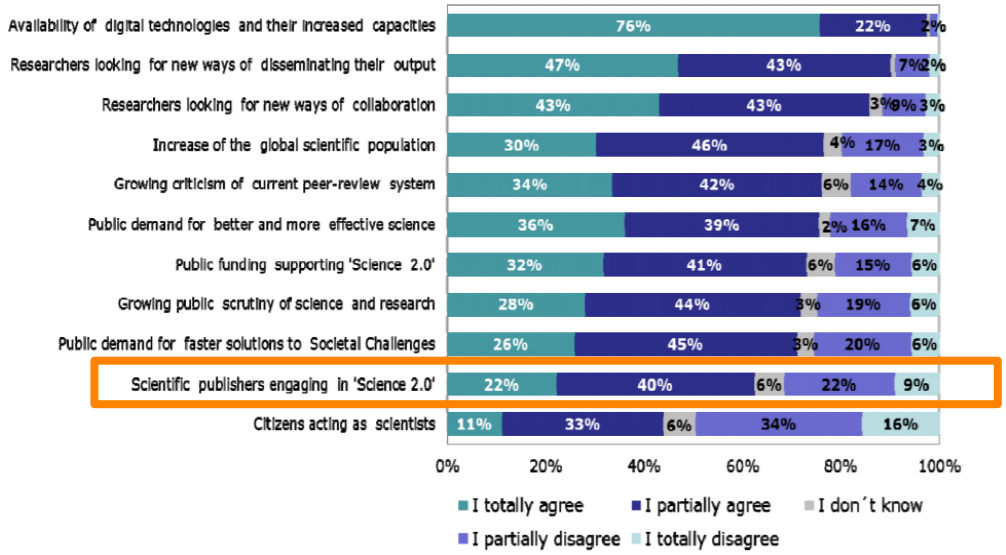
- Accessible
- Collaborative
- Transparent
- Effective
- Efficient

Through:

- Encouraging a culture of openness and sharing
- Leveraging and developing new technologies
- Developing and adapting reward and metric systems

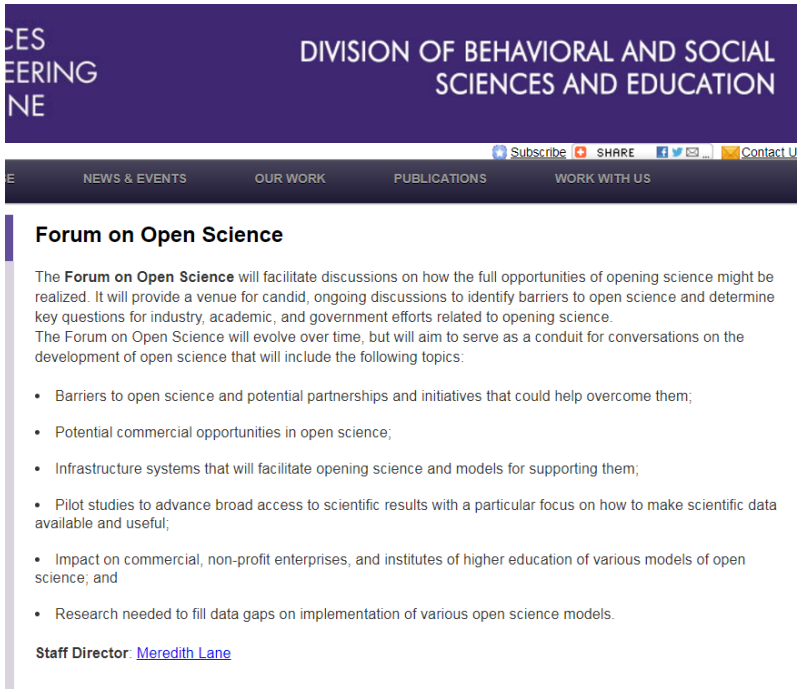
Publishers are a Key Stakeholder in Open Science

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



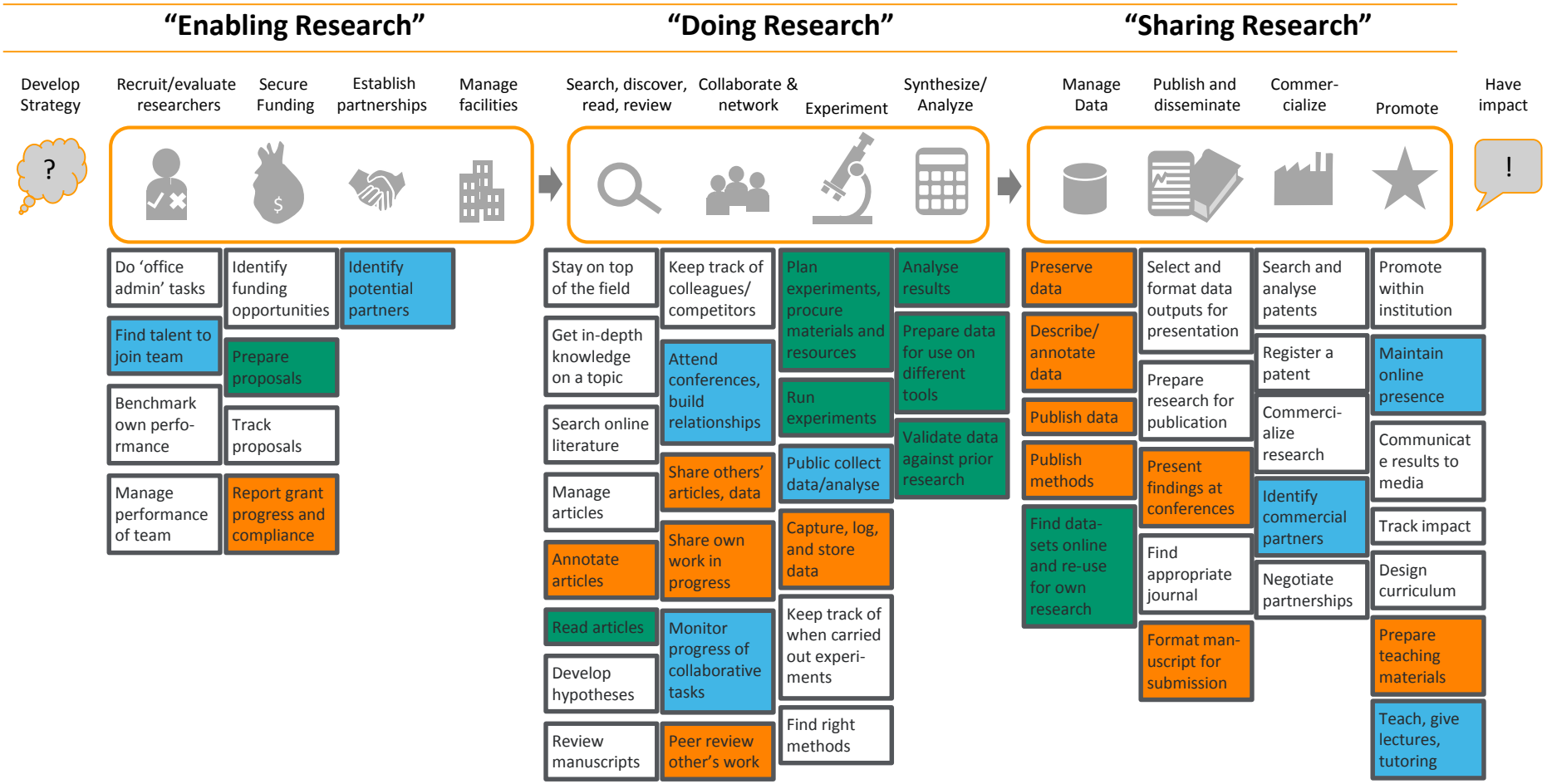
Sample size: 492, missing: 8 to 12.

http://ec.europa.eu/digital-agenda/en/news/final-report-science-20-public-consultation



http://sites.nationalacademies.org/DBASSE/CurrentProjects/DBASSE_087866

Open Science Across the Researcher Workflow



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

ELSEVIER

- **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

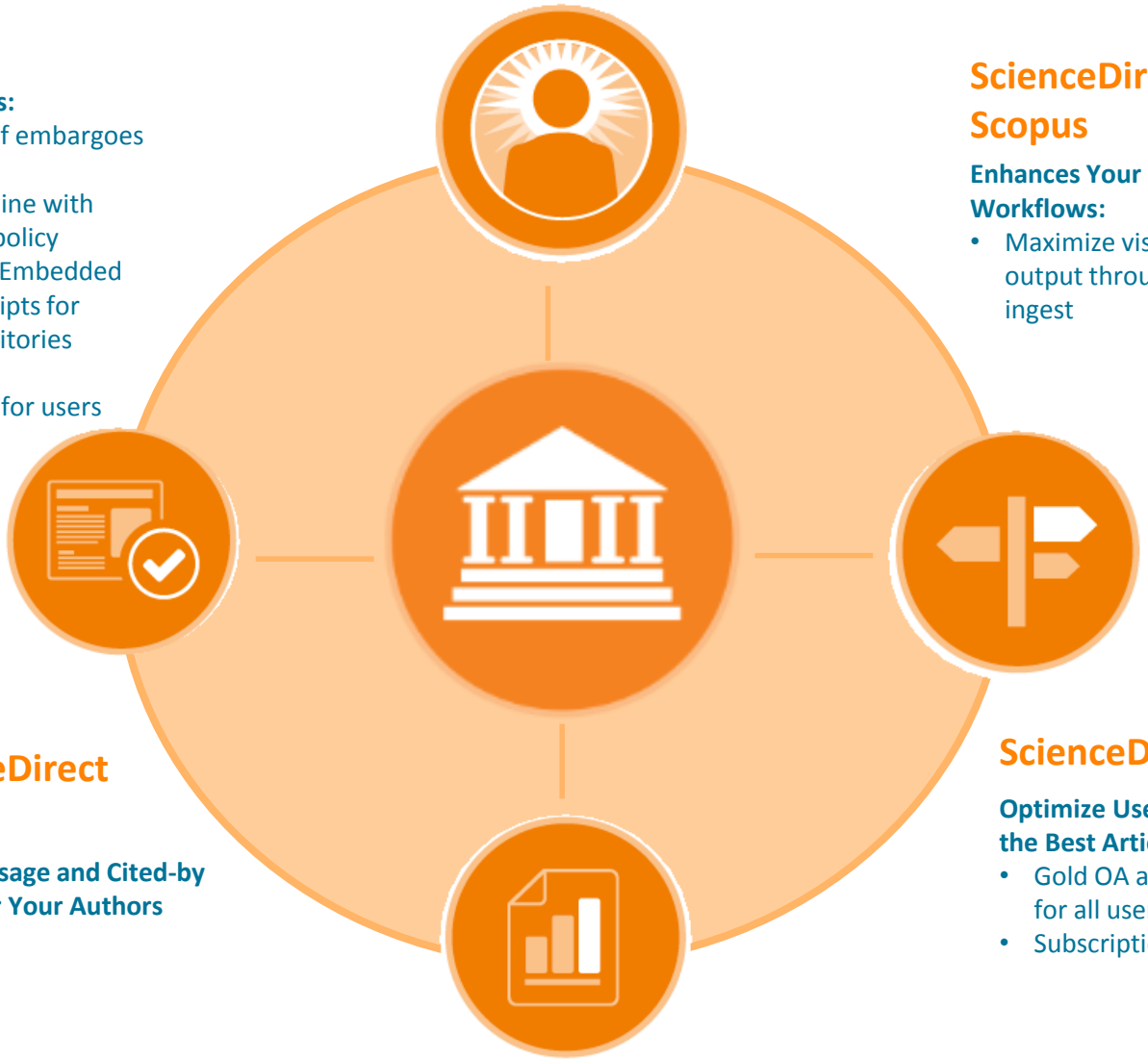
Public Access Services:

- 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

ScienceDirect Scopus

Enhances Your Reputation and Workflows:

- Maximize visibility of research output through automatic API ingest



ScienceDirect Scopus

Capture Usage and Cited-by Counts for Your Authors

ScienceDirect

Optimize User Experience/Guide to the Best Article Version:

- Gold OA and Accepted Manuscripts for all users
- Subscription articles for subscribers

www.elsevier.com/solutions/sciencedirect/support/institutional-repository

Institutio

ScienceDirect

Public Access Servi

- Automatic inges for AMs
- Article versions i sharing & hostin
- New Pilot Servic Accepted Manus Institutional Rep
- Service: includes downloadable A after embargo

ct

eputation and

bility of research
h automatic API

Scien
Scopu

Capture
Counts f

rect

Experience/Guide to
e Version:
d Accepted Manuscripts

n articles for subscribers

"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)

www.elsevier.com/solutions/sciencedirect/support/institutional-repository

Elsevier Partners with CHORUS

CHORUS

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

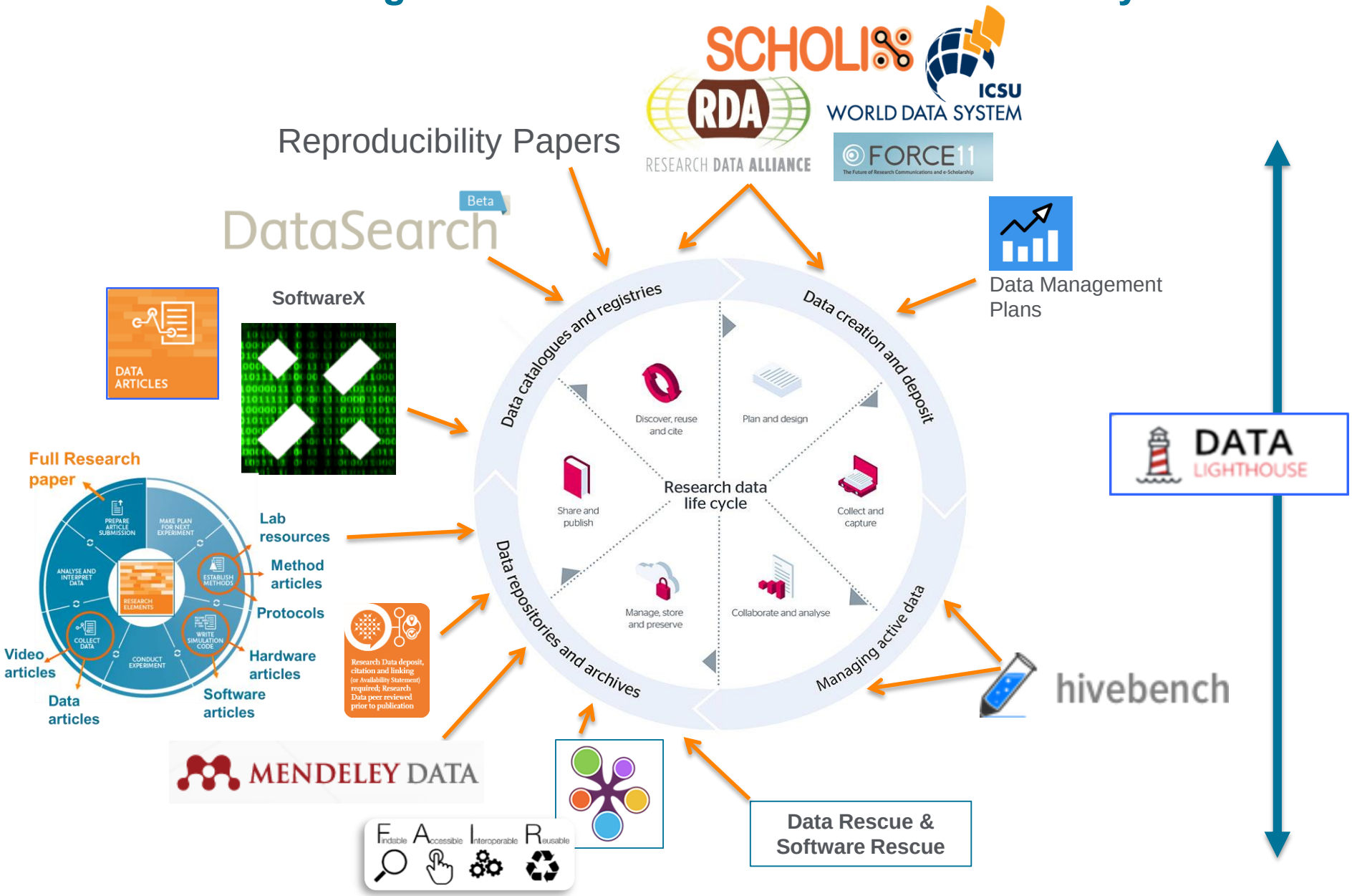
Why make it open?

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"

2017 DSA 2019

MENDELEY DATA Browse My datasets New dataset

You are previewing this dataset.

TB-Drugome - Run 126

Contributor(s): Sarah L. Kinnings, Li Xie, Kingston H.

Description of this data

A run of the TB-Drugome Workflow.

Experiment data files

IN:run_126_worflow.png	
IN:smap_properties.txt	
IN:drug_key.csv	
IN:drug_binding_sites_list.csv	
IN:MtbStructures_tar_gz.gz	
IN:homology_models_list_txt.txt	17 KB
IN:tb_protein_info.csv	42 KB
IN:template_pdbs.csv	24 KB

Describe how experiment can be reproduced

Steps to reproduce

The TB-Drugome Workflow is complicated to use without the proper knowledge in bioniformatics and computer science. This section aims to provide the initial pointers and references to look at when trying to reuse the TB-Drugome workflow.

- It is recommended to first understand the flow of the data by looking at published run schema.
- Then look at the original Drugome paper for further information and the tools being used.
- Once you are familiar with the workflow, have a look at the published workflow run, decompose the workflow in the different modules and test them separately.
- In order to test the different steps of the workflow, you can access http://wind.isi.edu/marbles/ and import the "Drugome" domain. Different workflows already encode the sub parts of the workflow and are ready to be run. Note that you will need to ask for a user and password (contact Yolanda Gil, Varun Ratnakar or Daniel Garijo).
- You will find suitable inputs to test the workflow in the input data section of this document. If you reuse the inputs, the outputs should look like the ones described here.

Latest version

Draft (of version 2)

You are previewing this dataset. You can share this preview privately or publish it to share it with the world.

Reserved DOI:
doi:10.17632/r69mvkckmn.2

Cite this dataset

Kinnings, Sarah L.; Xie, Li; Fung, Kingston H.; Jackson, Richard M.; Xie, Lei; Bourne, Philip E. (2016), "TB-Drugome - Run 126", Mendeley Data, v2

http://dx.doi.org/10.17632/r69mvkckmn.2

DOI is reserved but not active

Create DOI for Citation

Previous versions

Version 1 2016-09-09

Store up to 5 GB of data in many formats

Link back to protocols

Keep track of versions of dataset

Related links

https://www.hivebench.com/notebooks/8524/experiments/20562 entity is source of this 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

PLUMX

USAGE
(clicks, downloads, views, library holdings, video plays)

CAPTURES
(bookmarks, code forks, favorites, readers, watchers)

MENTIONS
(blog posts, comments, reviews, Wikipedia links)

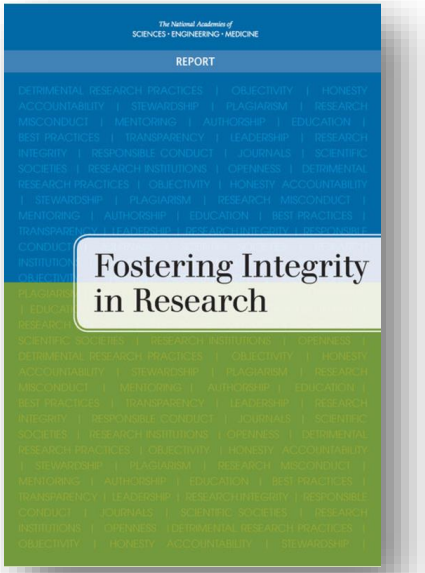
SOCIAL MEDIA
(+1s, likes, shares, tweets)

CITATIONS
(citation indexes, patent citations, clinical citations)

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

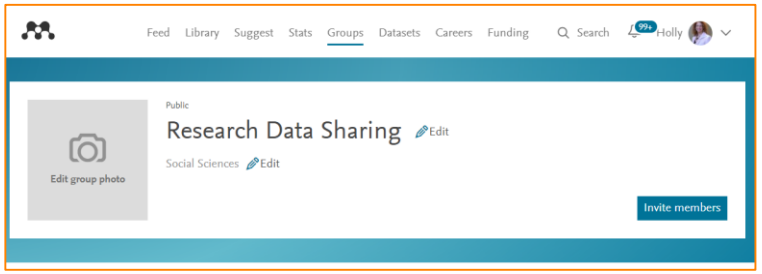
https://rdmi.uchicago.edu/papers/08212017144742_deWaard082117.pdf

Additional Resources



NASEM Consensus Study Report (2017), **BOX 9-3: Best Practices Checklist for Journals**

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.

"Advancing Open Science through Partnerships"

NFAIS CONFERENCE
Open Access & Beyond
Embassy Suites Old Town Alexandria | OCT 2-3, 2017

Additional Resources

Anderson, Rick. 2015. Open Access and Academic Freedom. Inside Higher Ed. December 15. https://www.insidehighered.com/views/2015/12/15/mandatory-open-access-publishing-can-impair-academic-freedom-essay

Davis PM. 2013. Public accessibility of biomedical articles from PubMed Central reduces journal readership—retrospective cohort analysis. *FASEB Journal* 27 (7): 2536-2541 http://dx.doi.org/10.1096/fj.13-229922.

Davis, PM. Journal Usage Half-Life. November 25, 2013. Association of American Publishers. http://www.publishers.org/_attachments/docs/journalusagehalflife.pdf

Husen, Sean, de Wilde, Zoe G., de Waard, Anita and Cousijn, Helena. Recommended Versus Certified Repositories: Mind the Gap. (June 23, 2017). Available at SSRN https://ssrn.com/abstract=3020994. The study's underlying research data is available at https://data.mendeley.com/datasets/zx2kcyvvwm/1.

Lavizzari, Carlo Scollo and Viljoen, René (2015). Open Access Licensing, A PRC Guide. http://publishingresearchconsortium.com/index.php/prc-guides-main-menu/166-open-access-licensing-0215

Nelson, Cary. 2013. Open Access and Academic Freedom. Inside Higher Ed. November 15. https://www.insidehighered.com/views/2013/11/15/essay-impact-open-access-requirements-academic-freedom

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 eContent*plus* 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.

STM Response to the Max Planck Digital Library White Paper on the Open Access Transition, http://www.stm-assoc.org/2015_12_04_STM_Response_to_MPDL_Open_Access_Transition_White_Paper.pdf

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.

ELSEVIER

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

Empowering Knowledge