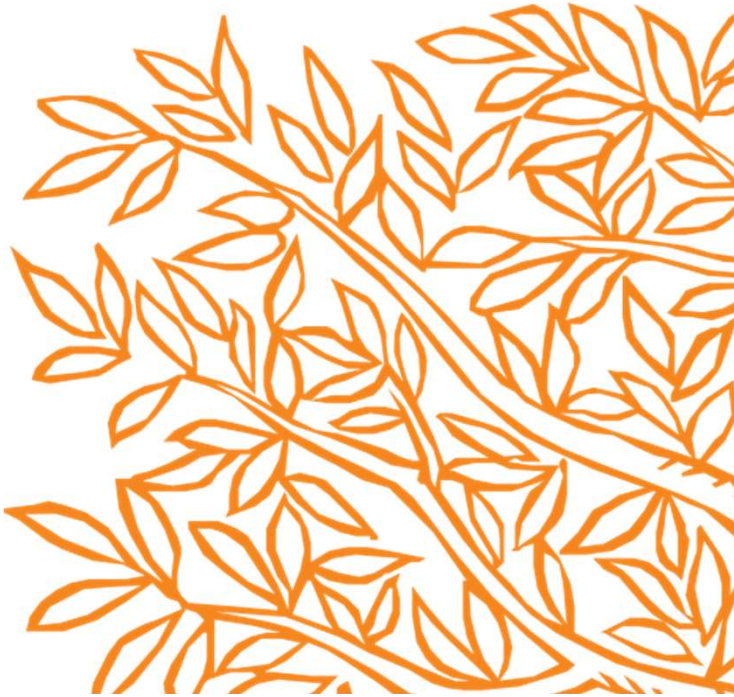
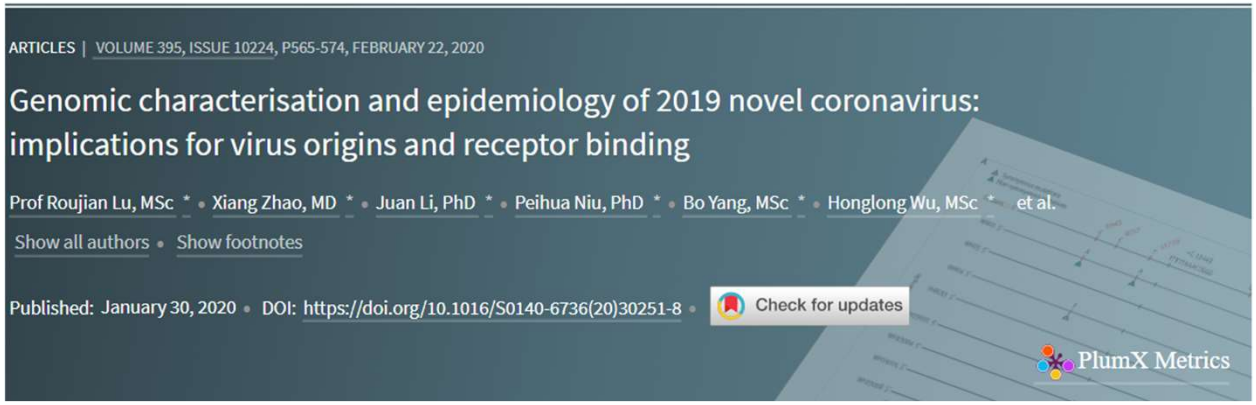
Open Science and Research Data Management - Trends and Practical Examples of Implementations

Dr. Anders Karlsson, Vice President, Global Strategic Networks, Elsevier
a.karlsson@elsevier.com

Adam Goh,
Regional Solutions Manager Software Solutions, Elsevier
adam.goh@elsevier.com



Early genomic characterization published on Jan 30 with Data available for interpretation



From the nine patients' samples analysed, eight complete and two partial genome sequences of 2019-nCoV were obtained. These data have been deposited in the China National Microbiological Data Center (accession number NMDC10013002 and genome accession numbers NMDC60013002-01 to NMDC60013002-10) and the data from BGI have been deposited in the China National GeneBank (accession numbers CNA0007332–35).

https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(20)30251-8/fulltext

Use LIGO Data
LIGO's Impact on Science and Technology
LIGO R&D
Detection Papers
Collaborate
All Publications
Observatory Status
The LIGO Laboratory Award for Excellence in Detector Characterization and Calibration

Use LIGO Data

The Gravitational-Wave Open Science Center

The Gravitational-Wave Open Science Center (GWOSC) fulfills LIGO's commitment to release, archive, and serve LIGO data to the broader scientific community and to the public, and to provide the information and tools necessary to understand and use the data. This resource is also a useful tool for teaching gravitational-wave data analysis to students around the world. The GWOSC site includes detailed metadata, tutorials, tools, and software to help public users perform effective analyses.

The LIGO Laboratory's Data Management Plan describes the scope and timing of LIGO data releases.

GW200105_162426 data release

GW200115_042309 data release

O3 bulk strain data release

GWTC-2 data release

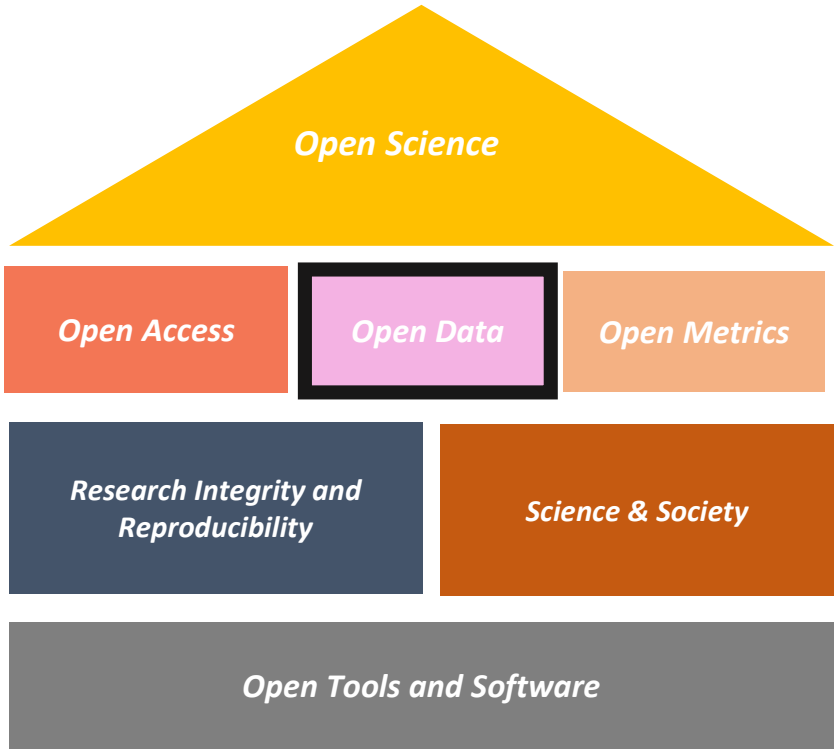
GW190521 data release

Content of presentation

Research Data Management as Part of Open Science and a tool to Advance Uptake and Impact of Scientific Knowledge
Dr. Anders Karlsson, Vice President, Global Strategic Networks, Elsevier

Practical RDM – Examples and Solutions from Elsevier
Adam Goh, Regional Solutions Manager Software Solutions, Elsevier

Open Science as an enabler for research excellence



Open Science describes a way of working which makes research more inclusive, more collaborative and more transparent

OS can benefit research and society and drive research performance by….

1. Enhancing knowledge sharing and collaboration beyond academia
2. Breaking down barriers to Reproducibility
3. Opening data for re-use and data-partnerships
4. Adopting open and transparent metrics for responsible research evaluation

ELSEVIER

Why should I share my data?

INCREASE IMPACT OF THE RESEARCH

IMPROVE QUALITY OF THE PUBLISHABLE RESEARCH OUTPUTS

MAKE RESULTS TRUSTWORTHY

IMPROVE EDUCATIONAL EXPERIENCE

INCREASE CITATIONS

INCREASE REPUTATION

GET ADDITIONAL PUBLICATION (DATA-IN-BRIEF)

Sharing research data leads to improved citations and academic reputation

The citation advantage of linking publications to research data

Giovanni Colavizza[1,2,*], Iain Hrynaszkiewicz[3,4], Isla Staden[1,5], Kirstie Whitaker[1,6], Barbara McGillivray[1,6]

1 The Alan Turing Institute, UK.
2 University of Amsterdam, NL.
3 Springer Nature, UK.
4 Public Library of Science, UK.
5 Queen Mary University, UK.
6 University of Cambridge, UK.

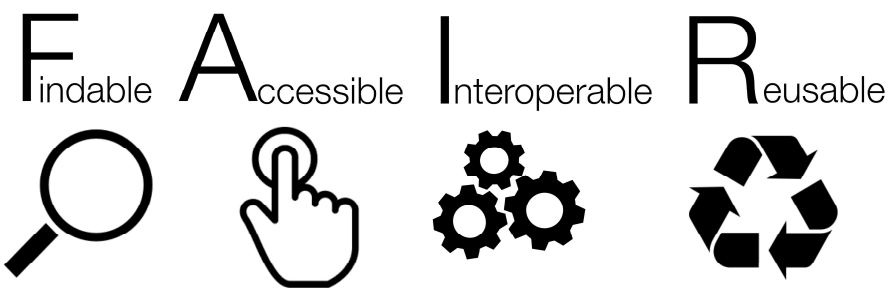
Abstract

Efforts to make research results open and reproducible are increasingly reflected by journal policies encouraging or mandating authors to provide data availability statements. As a consequence of this, there has been a strong uptake of data availability statements in recent literature. Nevertheless, it is still unclear what proportion of these statements actually contain well-formed links to data, for example via a URL or permanent identifier, and if there is an added value in providing such links. We consider 531, 889 journal articles published by PLOS and BMC, develop an automatic system for labelling their data availability statements according to four categories based on their content and the type of data availability they display, and finally analyze the citation advantage of different statement categories via regression. We find that, following mandated publisher policies, data availability statements become very common. In 2018 93.7% of 21,793 PLOS articles and 88.2% of 31,956 BMC articles had data availability statements. Data availability statements containing a link to data in a repository—rather than being available on request or included as supporting information files—are a fraction of the total. In 2017 and 2018, 20.8% of PLOS publications and 12.2% of BMC publications provided DAS containing a link to data in a repository. We also find an association between articles that include statements that link to data in a repository and up to 25.36% (± 1.07%) higher citation impact on average, using a citation prediction model. We discuss the potential implications of these results for authors (researchers) and journal publishers who make the effort of sharing their data in repositories. All our data and code are made available in order to reproduce and extend our results.

Sharing data works
25.36%
higher citation impact

Source: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0230416 (Published April 2020)

FAIR principles for research data management



SCIENTIFIC DATA

Comment | OPEN | Published: 15 March 2016

The FAIR Guiding Principles for scientific data management and stewardship

Mark D. Wilkinson, Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, Philip E. Bourne, Jildau Bouwman, Anthony J. Brookes, Tim Clark, Mercè Crosas, Ingrid Dillo, Olivier Dumon, Scott Edmunds, Chris T. Evelo, Richard Finkers, Alejandra Gonzalez-Beltran, Alasdair J.G. Gray, Paul Groth, Carole Goble, Jeffrey S. Grethe, Jaap Heringa, Peter A.C 't Hoen, Rob Hooft, Tobias Kuhn, Ruben Kok, Joost Kok, Scott J. Lusher, Maryann E. Martone, Albert Mons, Abel L. Packer, Bengt Persson, Philippe Rocca-Serra, Marco Roos, Rene van Schaik, Susanna-Assunta Sansone, Erik Schultes, Thierry Sengstag, Ted Slater, George Strawn, Morris A. Swertz, Mark Thompson, Johan van der Lei, Erik van Mulligen, Jan Velterop, Andra Waagmeester, Peter Wittenburg, Katherine Wolstencroft, Jun Zhao & Barend Mons - Show fewer authors

DOI: 10.1038/sdata.2016.18

FAIR Principles

GO FAIR is committed to making data and services **findable**, **accessible**, **interoperable** and **reusable (FAIR)**.

Findable: Metadata and data should be easy to find for both humans and computers.

Accessible: The exact conditions under which the data is accessible should be provided in such a way that humans and machines can understand them.

Interoperable: The (meta)data should be based on standardized vocabularies, ontologies, thesauri etc. so that it integrates with existing applications or workflows.

Reusable: Metadata and data should be well-described so that they can be replicated and/or combined in different research settings.

https://www.go-fair.org/

Data sharing is becoming standard practice

Three trends are coming together,

1) Computational:

- Faster, easier, cheaper, more computational methods of doing science
- Coming of age of analytics yield new layers of insight on same data

2) Funding bodies & institutions:

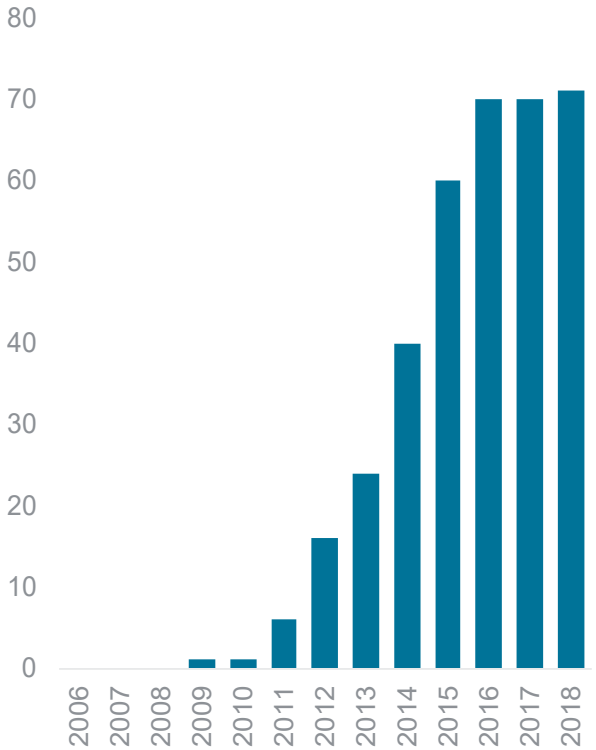
- Funding agencies driving data sharing mandates aiming to improve re-use of data and reproducibility of research
- Which is followed by institutions adopting data management policies and plans

3) Researchers:

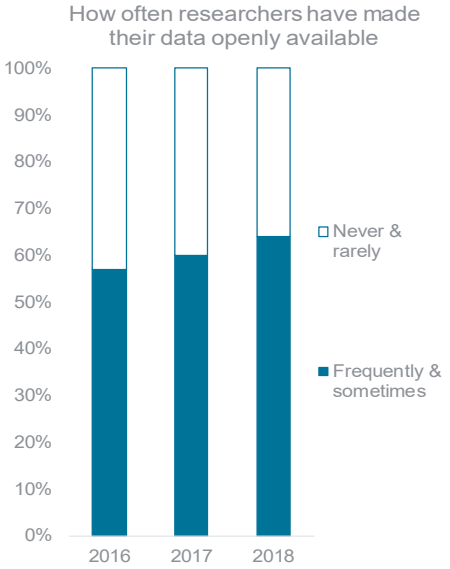
- a new generation of researchers, more focused on data and data sharing
- "Research Data is a first-class citizen"

Example of trend 2

Number of UK institutions adopting a data management policy



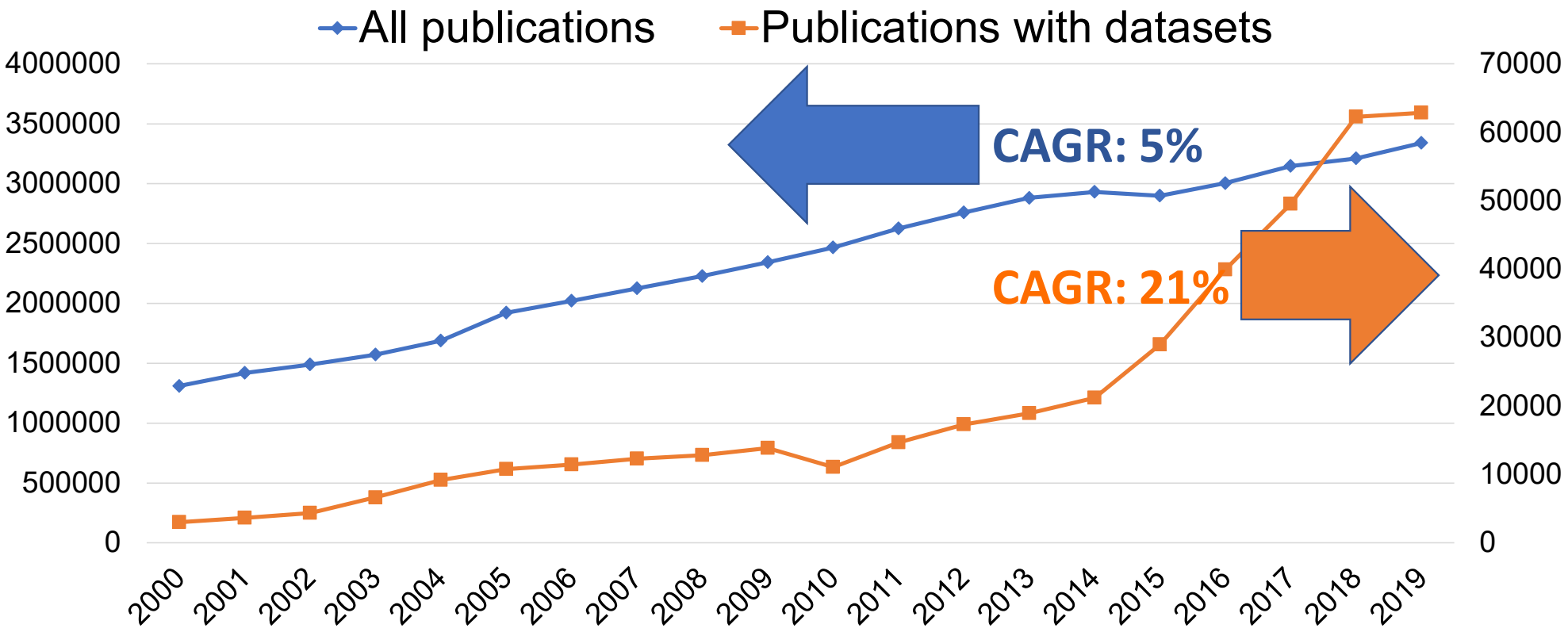
Examples of trend 3:



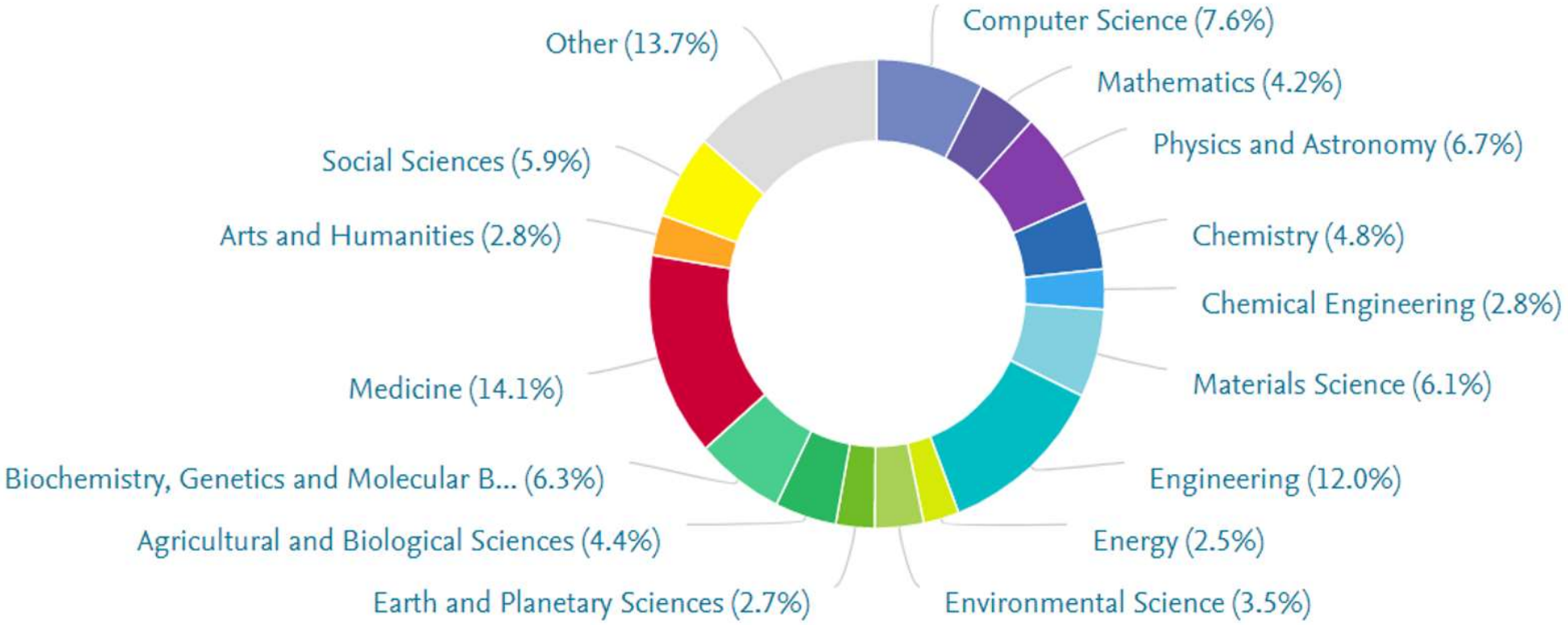
How researchers value data citations		
Date of first peer reviewed publication	Respondents who value a data citation same or more than an article citation	Respondents who value a data citation less than an article citation
<1990	64%	36%
90s	65%	35%
00s	60%	40%
10s	72%	28%

Source: http://www.dcc.ac.uk/resources/policy-and-legal/institutional-data-policies
http://v2.sherpa.ac.uk/juliet/
https://figshare.com/articles/The_State_of_Open_Data_Report_2018/7195058 (normalized per year bracket)

Research Data Management Adoption is Growing Worldwide
(note different scales on left and right axis)



Publications with datasets globally –(2015-2019)



www.elsevier.com/about/open-science

Transparent research

Breaking down barriers to reproducibility: the challenges

- No common definition across all fields
- Lack of ***incentives*** to conduct replications
- Researchers ambivalent about ***sharing data***
- ***Methods*** lacking in transparency/rigor
- Difficult to ***validate*** reproducibility during peer review
- Insufficient rigour in ***statistics***

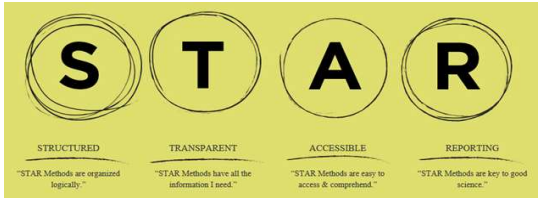
Data in Brief

MethodsX

HardwareX

SoftwareX

Software Impacts



Research Elements journals

The Research Elements journals are multidisciplinary, open access peer-reviewed journals that publish digestible articles on research objects. Research Elements articles may *complement* an original research article or they may be *independent*.

Data in Brief

SoftwareX

Invention Disclosure

MethodsX

Software Impacts

STAR Protocols

HardwareX

Research article & Data article

Journal of Colloid and Interface Science
Volume 581, Part B, 1 January 2021, Pages 644-655

Green earth pigments dispersions: Water dynamics at the interfaces

Agathe Fanost [a, b], Maguy Jaber [b], Laurence de Viguerie [b], Jean-Pierre Korb [a], Pierre E. Levitz [a], Laurent J. Michot [a], Guillaume Mériguet [a], Anne-Laure Rollet [a]

Show more

https://doi.org/10.1016/j.jcis.2020.07.085

Get rights and content

Referred to by Agathe Fanost, Maguy Jaber, Laurence de Viguerie, Jean-Pierre Korb, Pierre E. Levitz, Laurent J. Michot, Guillaume Mériguet, Anne-Laure Rollet

Green Earth pigments aqueous dispersions: NMR relaxation rates dataset

Data in Brief, Volume 32, October 2020, Pages 106270

Data in Brief
Volume 32, October 2020, 106270

Data Article

Green Earth pigments aqueous dispersions: NMR relaxation rates dataset

Agathe Fanost [a, b], Maguy Jaber [b], Laurence de Viguerie [b], Jean-Pierre Korb [a], Pierre E. Levitz [a], Laurent J. Michot [a], Guillaume Mériguet [a], Anne-Laure Rollet [a]

Show more

https://doi.org/10.1016/j.dib.2020.106270

Get rights and content

open access

Refers to Agathe Fanost, Maguy Jaber, Laurence de Viguerie, Jean-Pierre Korb, Pierre E. Levitz, Laurent J. Michot, Guillaume Mériguet, Anne-Laure Rollet

Green earth pigments dispersions: Water dynamics at the interfaces

Journal of Colloid and Interface Science, Volume 581, Part B, 1 January 2021, Pages 644-655

Journal data sharing policies at Elsevier

Elsevier journals data sharing policies

- Rolled out in 2017
- Default B suggested to proprietary titles, editors opt out/decide on different level.
  Most journals (1800+ in B, 75 journals in policy C, a few in policy D, a good number A/opt out)
- For journal to support FAIR Data: policy C and above

https://www.elsevier.com/authors/author-services/research-data/data-guidelines

Cell Press journal suite

- Own policies: data archiving strong recommendation
- STAR Methods: *Structured, Transparent, Accessible Reporting*

TOP GUIDELINES
TRANSPARENCY AND OPENNESS PROMOTION

https://www.cos.io/initiatives/top-guidelines

The Lancet journal suite

- Requires *Data sharing statement* for all research articles since Sep 2020
- First ICMJE member to expand on 2018 data sharing statement for clinical trials data

Elsevier contribution to the research data community

Force11

Co-founder
Co-author FAIR Data principles
Implementation data citations principles

ORCID

Co-founder

Pistoia Alliance

Active member

ICSU

Active member

Scholix

Co-founder

STM

Supporting Brussels open data declaration

Research Data Alliance

Active member

https://www.elsevier.com/open-science/research-data

www.elsevier.com/about/open-science

Stick or carrot? To summarize my part

Three trends are coming together, leading to an acceleration of RDM practices

1) Computational:

- Faster, easier, cheaper, more computational methods of doing science
- Coming of age of analytics yield new layers of insight on same data

2) Funding bodies & institutions:

- Funding agencies driving data sharing mandates aiming to improve re-use of data and reproducibility of research
- Which is followed by institutions adopting data management policies and plans

3) Researchers:

- a new generation of researchers, more focused on data and data sharing
- “Research Data is a first-class citizen”

How Elsevier makes RDM easier

Practical RDM – Examples and Solutions from Elsevier

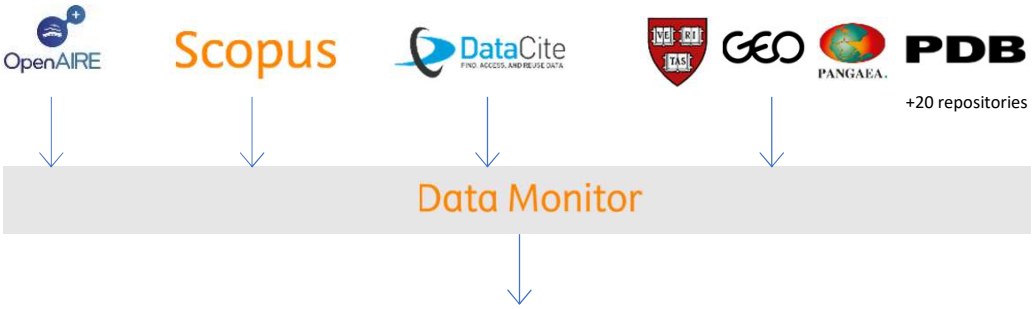
Adam Goh, Regional Solutions Manager Software Solutions, Elsevier

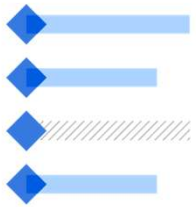
- RDM is more than data policies and data management plans
- RDM is about **helping** researchers and institutions with their data

Image source: https://wonkhe.com/blogs/too-much-information-not-enough-advice/

Integrating with industry-leading research tools

- Connects institutions to ~30 million research data records from 2000+ generalist and domain-specific repositories
- Integrates with services used by researchers and institutions worldwide





~30 mil

data records

2000+

domain repositories

Simplifying research data visibility

Actively monitoring thousands of data repositories researchers use often presents a challenges for institutions.

Our solution

Provides accurate visibility into the entire research data output of institutions

Integrates ~30 million research data records from 2000+ generalist and domain specific repositories

Research data repositories	Number of datasets
Zenodo	1817618
USGS Mineral Resources	995208
Cambridge Structural Database	909175
GEOROC	478309
PANGAEA	403887
RCSB-PDB	138270

Above: Example of repositories from Data Monitor corpus as of April 2021.

Source: https://datascience.codata.org/articles/10.5334/dsj-2017-042/

Datasets picked up by Data Monitor

3746 datasets from

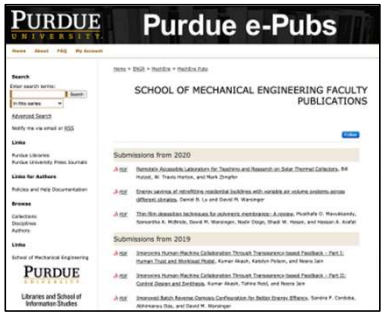
- The Cambridge Structural Database
- Zenodo
- HEPData
- Dryad
- And more…

Digital Commons Data, powered by Digital Commons

Embed your RDM program in your institution's larger strategy for Open Science and scholarly communications. A modular and extensible digital platform for all the institution's research output, Digital Commons helps institutions increase visibility and build a global reputation by showcasing the full spectrum of your scholarly output.

Highlight Faculty Publications *DC Repository*

Create a showcase out of your researcher's published and pre-published works



Manage Research Data *DC Data*

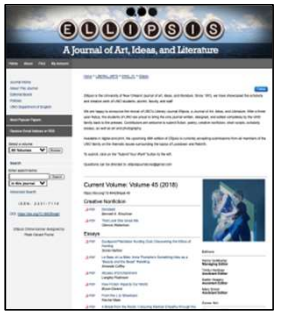
Drive forward your Research Data Management program, with specialized tools

Curate Digital Scholarship *DC Exhibits*

Curate exhibits of research projects and portfolios of works

Elevate Diamond OA Journals *DC Publishing*

Launch and develop an OA journal publishing program



The Digital Commons community surpassed one billion total downloads in year 2020

Report on dataset usage and engagement

Your institution's homepage will **display datasets associated to your institution,** powered by **Data Monitor,** which indexes 28+ million datasets from 2000+ domain-specific and generalist repositories.

In addition, you can **customize the content on your homepage,** by organizing datasets into collections. For example, departments, research themes or initiatives can all have their own collections featured on your homepage.

Institutional homepages host featured departmental collections. In addition, the institution's datasets are surfaced via advanced search.

ELSEVIER

Thank you