The reproducibility challenge – what researchers need

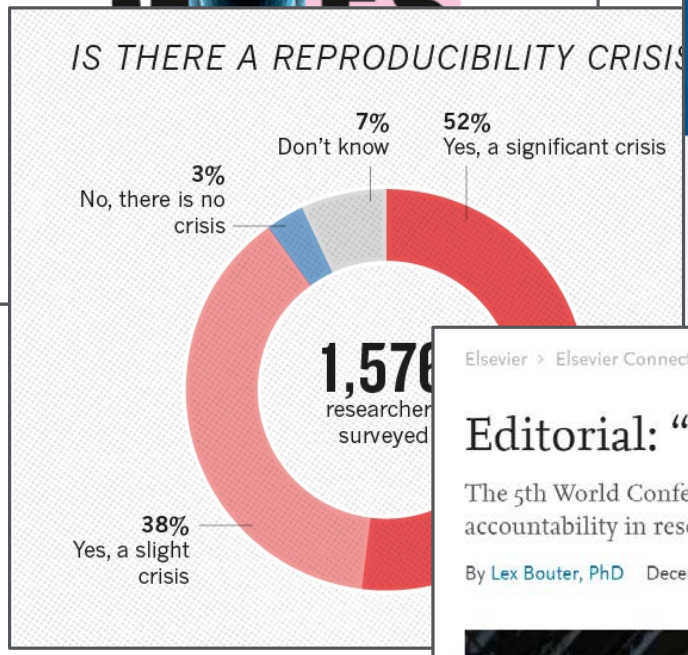
Federica Rosetta
Director Global Strategic Networks, Europe
ELSEVIER

Tromso, Munin Conference, 23-11-2017


The Economist
Washington's lawyer surplus
How to do a nuclear deal with Iran
Investment tips from Nobel economists
Junk bonds are back
The meaning of Sachin Tendulkar
HOW SCIENCE GOES


USA TODAY
Search
SUBSCRIBE NOW
3 MONTHS FOR $25
Press F11 to exit full screen
NEWS SPORTS LIFE MONEY TECH TRAVEL OPINION CROSSWORDS WASHINGTON VIDEO MORE
Column: Does science have honesty problem?
Alex Berezow and Hank Campbell Published 5:51 p.m. ET Nov. 28, 2012 | Updated 12:51 p.m. ET Nov. 29, 2012
New study shows an increase in retractions of scientific papers,

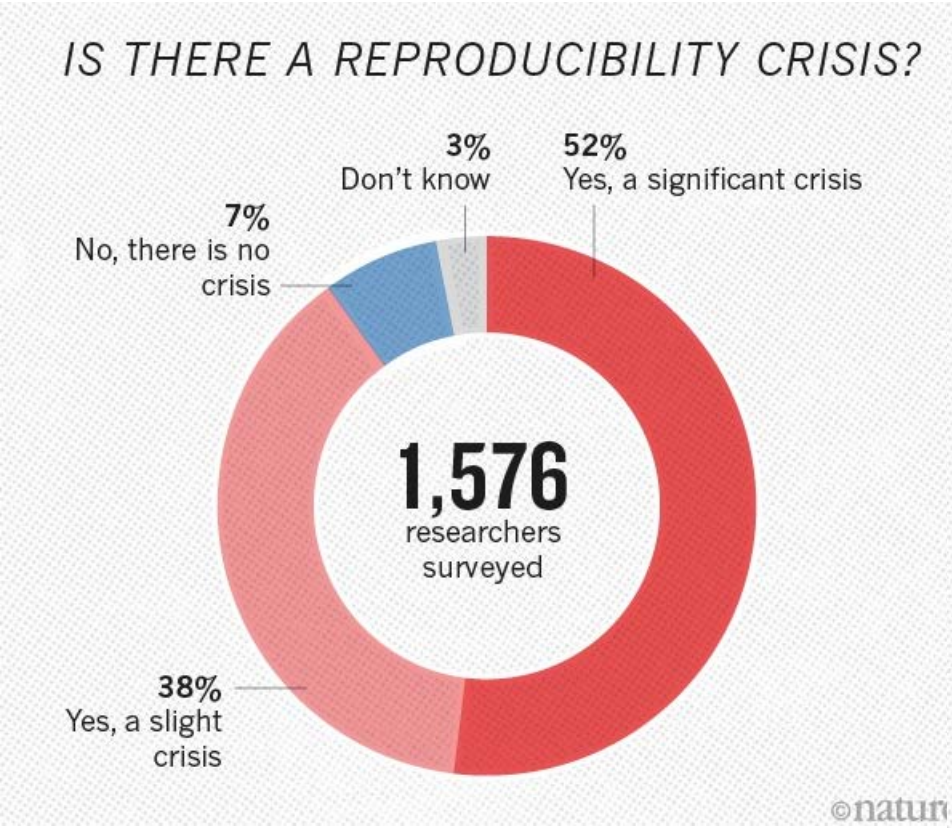

HOME
SEARCH
The New York Times
SCIENCE
Science, Now Under Scrutiny Itself

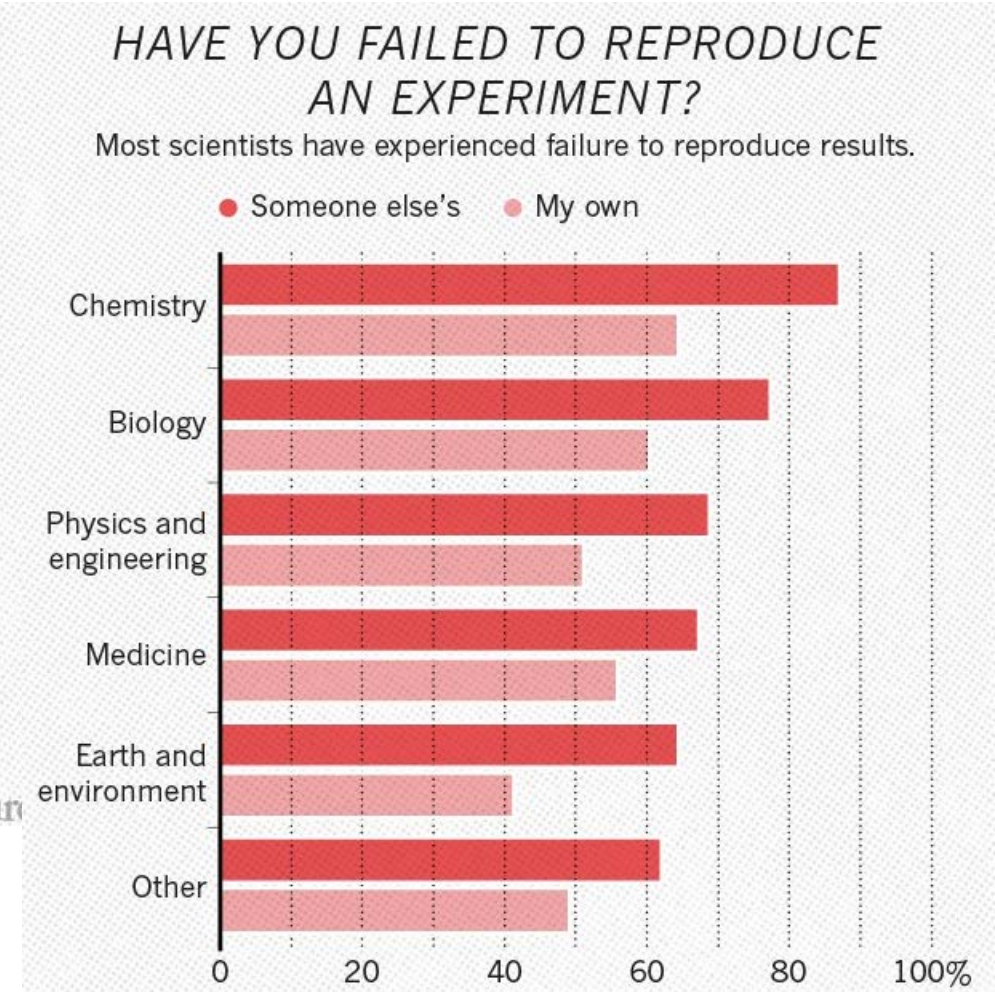

IS THERE A REPRODUCIBILITY CRISIS
7%
Don't know
52%
Yes, a significant crisis
3%
No, there is no crisis
38%
Yes, a slight crisis
researcher
surveyed


theguardian
website of the year
home › world
europe US americas asia australia africa middle east
all
South Korea
Disgraced cloning expert convicted for embezzlement and false claims


Elsevier › Elsevier Connect › Editorial: "Is science i...
Editorial: "Is science in big trouble?"
The 5th World Conference on Research Integrity will explore the challenges of promoting transparency and accountability in research — and develop an agenda for action
By Lex Bouter, PhD December 5, 2016

Researcher perception about Reproducibility



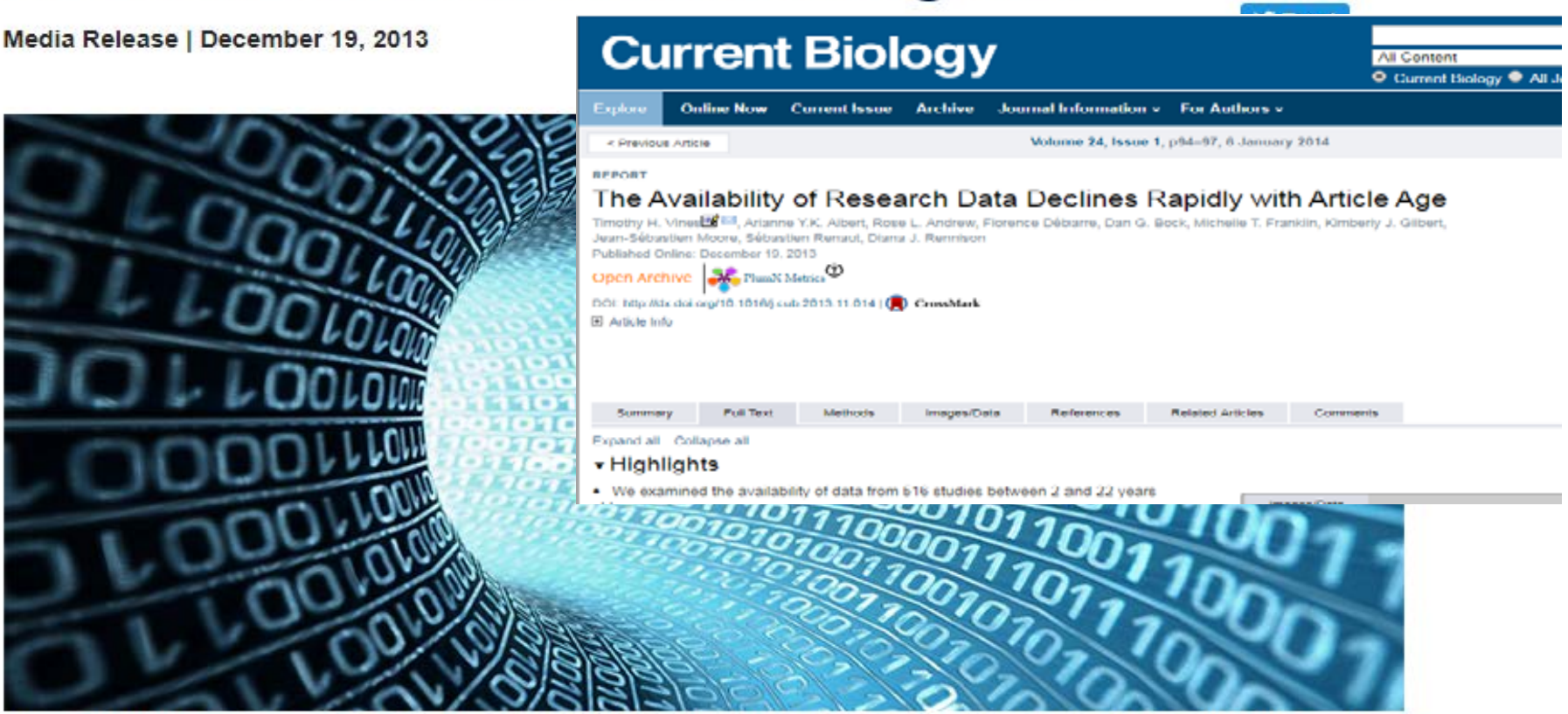


Ref: Nature 533 (2016)

Scientific data lost at alarming rate

Email

Media Release | December 19, 2013



A huge amount of scientific data is lost over time due to outdated e-mail addresses and storage devices, says UBC's Tim Vines. Photo: Mike_Kiev, iStock

Eighty per cent of scientific data are lost within two decades, according to a new study that tracks the accessibility of data over time.

The culprits? Old e-mail addresses and obsolete storage devices.

While all data sets were available two years after publication, the odds of obtaining the underlying data dropped by 17 per cent per year after that, they reported

Data sharing is important for science and society

nature International weekly journal of science

Home | News & Comment | Research | Careers & Jobs | Current Issue | Archive | Audio & Video | For A

New

NATU

Pa

High

anno

THE LANCET

Volume 383, Issue 9913, 18–24 January 2014, Pages 257–266

Ser

In

re

Dr

Pro

Ghe

Attention

a large m

Schizoph

ELSEVIER

Schizophrenia Research

Volume 163, Issues 1–3, April 2015, Pages 38–46

OPINION

The delay in sharing research data is costing lives

Josh Sommer

It is not uncommon for potentially life-saving research data to be published years after being generated. But the setback to progress caused by the delay in releasing data is troublesome for people who selflessly participate in trials and desperately await new therapies. Scientists need to feel greater urgency to share their findings quickly, and they need additional avenues to facilitate this process.

FOSTER About Resources Events Courses News Search for... Search

Definition of Open Reproducible Research

The act of practicing Open Science and the provision of offering to users free access to experimental elements for research reproduction.

FOREWORD BY COMMISSIONER CARLOS MOEDAS

I am delighted to preface the first report of the Commission High Level Expert Group European Open Science Cloud. Upon set-up of the group, back in September 2015, I tasked its members to explore what was then only an embryonic policy idea: the European Open Science Cloud, a future infrastructure to support Open Research Data and Open Science in Europe. Since then, the Commission made the idea a vision for the future of open science, in the context of the Digital Single Market.

Open Science is indeed one of my priorities. It is a move towards better science, to get more value out of our investment in science and to make research more reproducible and transparent. The scientific craft has changed beyond recognition in the last decade. Long gone is the era of the lonely scientist in the lab. Recent advances such as the discovery of the Higgs boson and gravitational waves, decoding of complex genetic schemas, climate change models, all required thousands of scientists to collaborate on ideas and, crucially, on data. And that implies that research data are findable and accessible and that they are interoperable and re-usable. In essence, this is what the Open Science Cloud is about: an open and trusted environment where research data can be safely stored and made openly available.

Ways to support reproducibility

ELSEVIER

Empowering Knowledge

When research starts: Registered Reports

- To reduce bias against negative results, to avoid changing study parameters after data collection (incl. prevention of p-hacking), to improve reproducibility, and to reward great ideas.
- Authors pre-register hypothesis & methods prior to doing research.
- If pre-registration accepted and research executed as registered, article with results will be published (also with negative results).
- 43 journals use the Registered Reports publishing format.
- Requires collaboration of journals and appreciation by funders.

Funders, associations, and institutes increasingly require data sharing

RESEARCH COUNCILS UK

Accessibility | Media Enquiries

search Search

Home | Funding | Research | Innovation | Skills | Public Engagement | News, Events and Publications | About Us

Home / Research / RCUK Common Principles on Data Policy

RCUK Common Principles on Data Policy

Making research data available to users is a core part of the Research Councils' remit and is undertaken in a variety of ways. We are committed to transparency and to a coherent approach across the research base. These RCUK common principles on data policy provide an overarching framework for individual Research Council policies on data policy.

Principles

NWO

contact | calendar | press | vacancies | Neder

Home | News & events | Funding | Research & results

Data management protocol

‹ Open Science

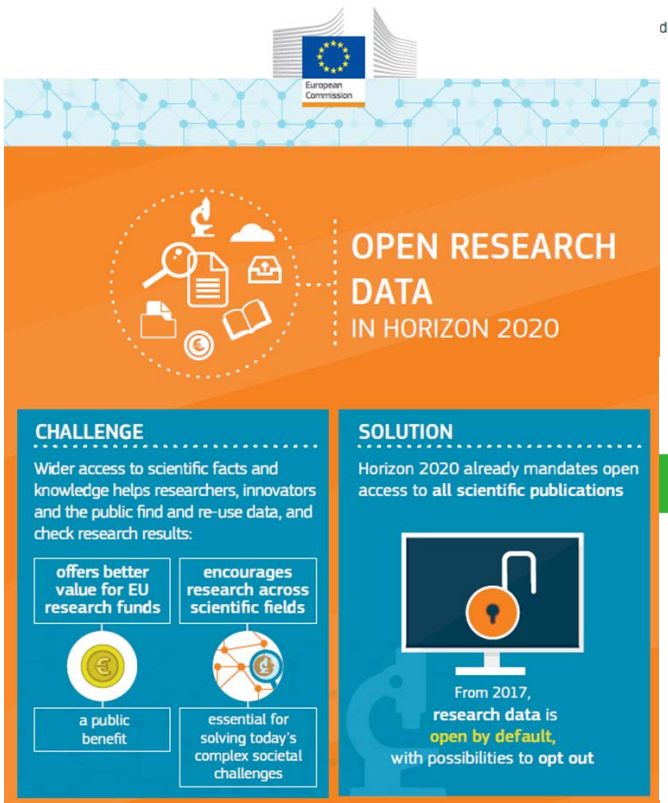
› Q&A Open Access at NWO

› Researchers about Open Access

Responsible data management is part of good research. NWO wants research data that emerges from publicly funded research to become freely and sustainably available, as much as possible, for the use by other researchers (data management as part of

d be made oper

European Commission



VSNU calls for a national RDM strategy

06 MAY 2015

The Association of Universities in the Netherlands (VSNU) has asked SURF to set up a National Research Data Management Coordination Point. The VSNU's Research and Valorisation Steering Group notes that much work in the RDM field is already taking place at universities.

vsnu

WAGENINGEN
UNIVERSITY & RESEARCH

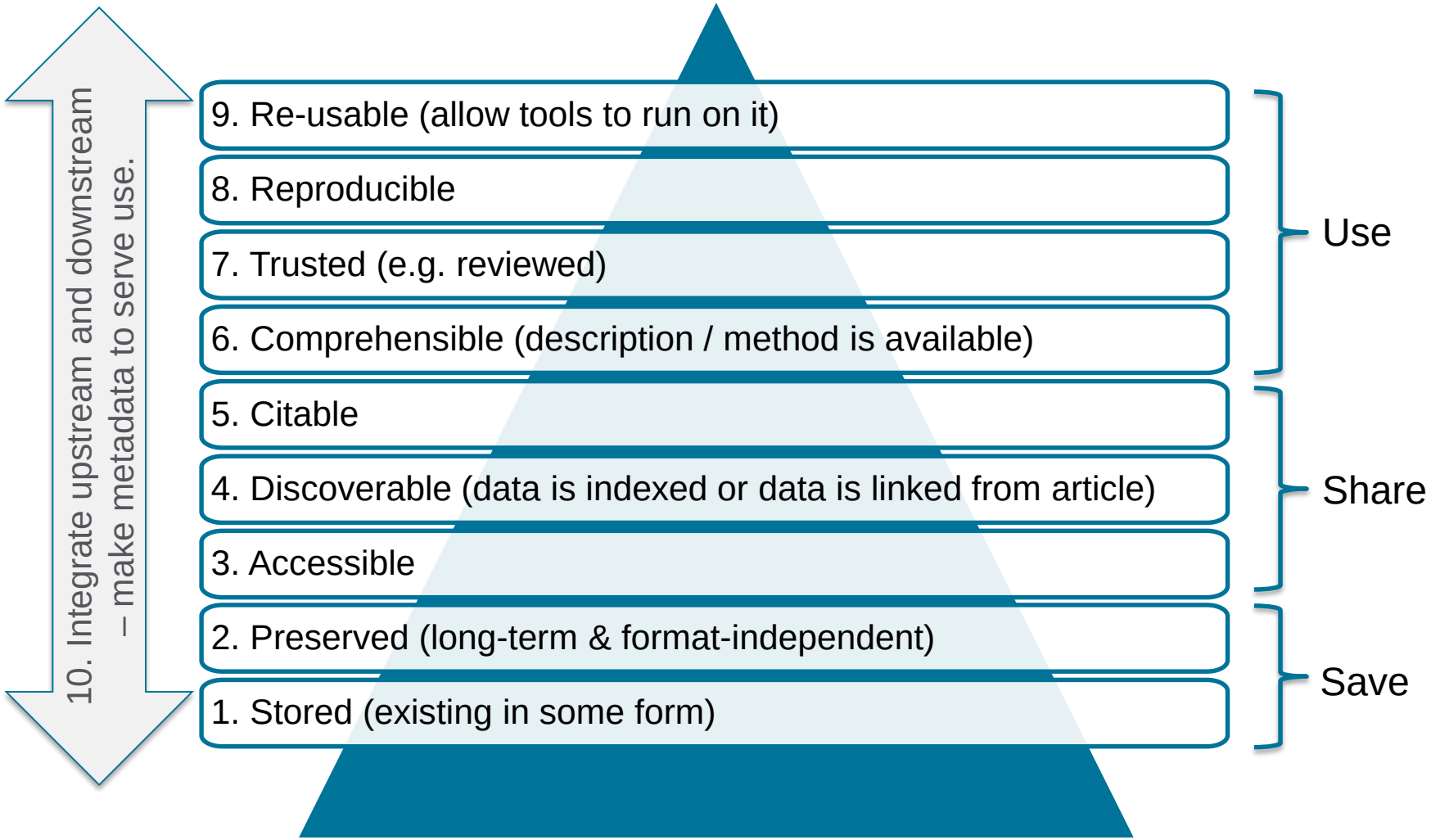
Home | Expertise & Services | Data Management Support Hub

Data Management Planning

Data management planning is a matter of good research practice. At Wageningen University & Research PhD candidates and Chair Groups are required to have a Data Management Plan.

National Coordination Point Research Data Management

While research is executed: Data Preservation

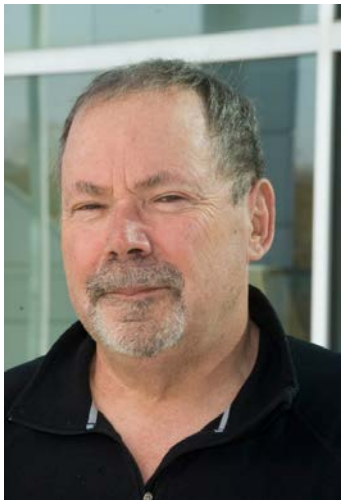


https://www.elsevier.com/connect/10-aspects-of-highly-effective-research-data

Old vs New Ways of Reproducing Research – high level summary of a real-life example

Take a standard experimental setup: e.g. tuberculosis research

- Phil Bourne, 2010 Investigated 'drugome' for tuberculosis, using computational scripts.



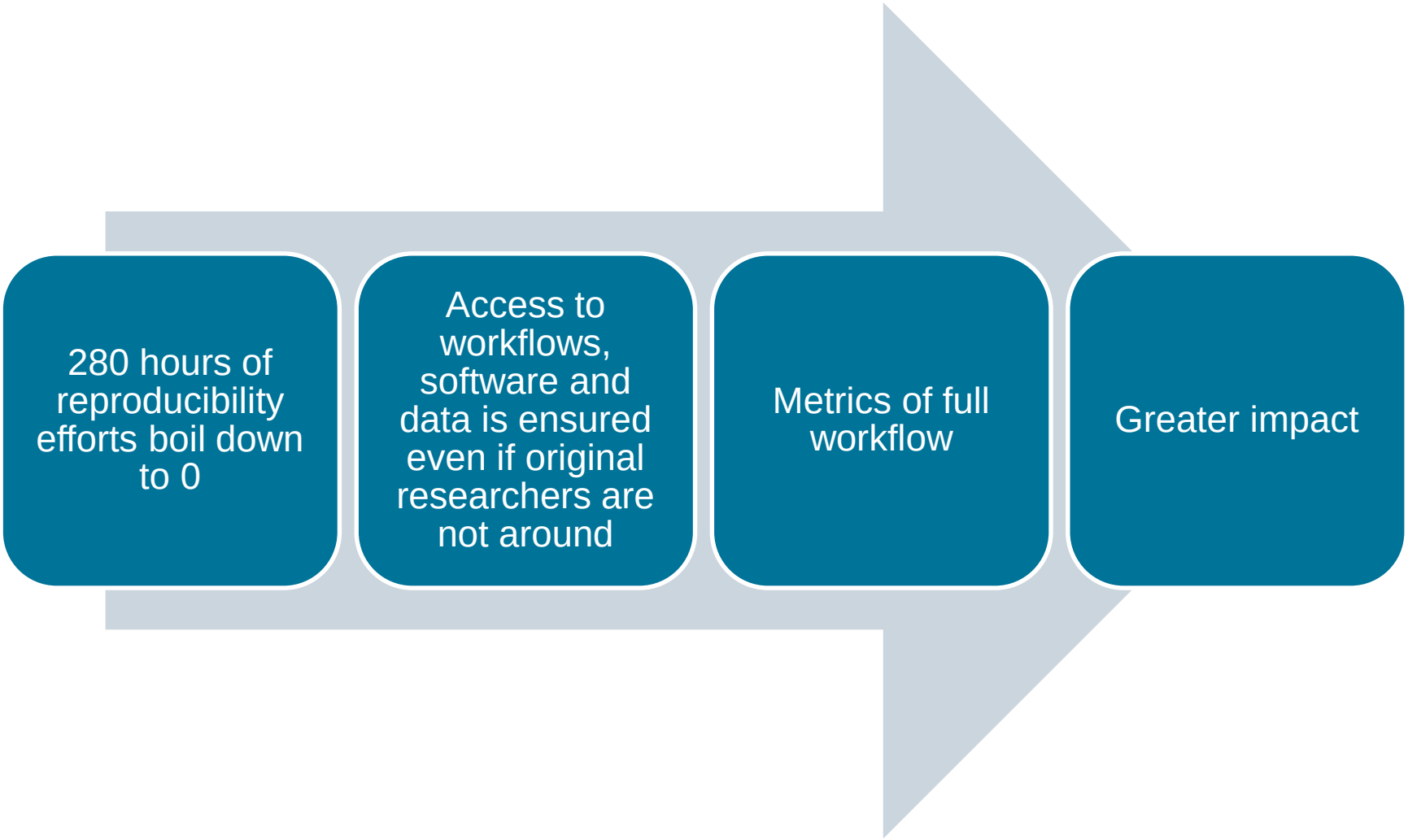
Key question: how much work is it to re-use or reproduce this experiment?

- Phil Bourne published a challenge to reproduce this in 2011; Yolanda Gil took up the challenge in 2012
- Spent280 hrs interviewing researchers, rerunning scripts, fixing broken code
- Experiment could be reproduced, but required:
  - A lot of work (280hours)
  - Full access to the original researchers

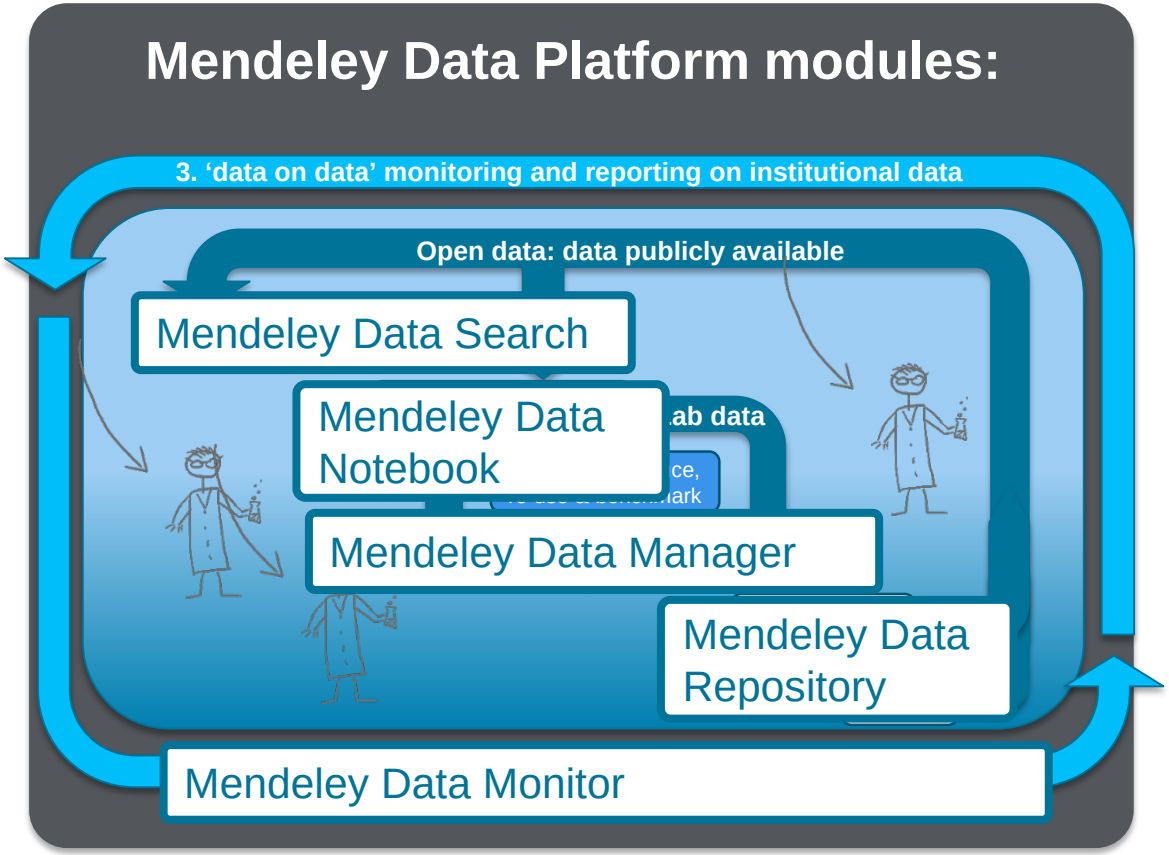
Challenge today: can we do this better, with less work, and without access to the original researchers?

- Answer: this is indeed possible (and was accomplished) if the researchers would have done their work within the same workflow, using an open ecosystem of interconnected tools
- Tools are not the answer alone: interconnectivity is
- Adoption is also a challenge, but can be overcome
- The toolset that was ultimately used in this example was Elsevier's Mendeley Data Platform

The results: cost-effective reproducibility, greater impact



The Mendeley Data Platform Modules to manage the entire lifecycle of research data



At Manuscript submission:
Comprehensive Journal Data Guidelines

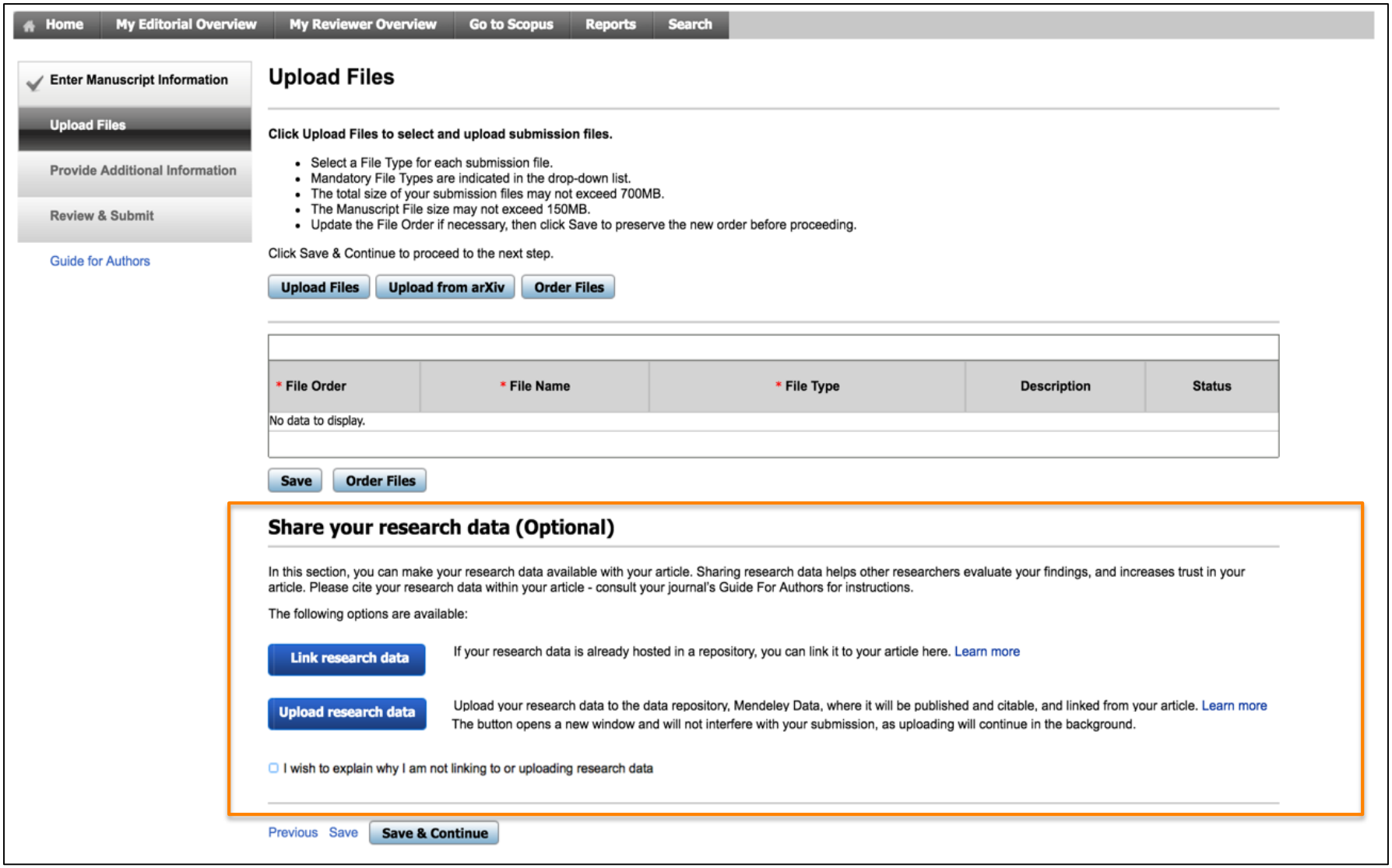
https://www.elsevier.com/authors/author-services/research-data/data-guidelines

https://cos.io/our-services/top-guidelines/

5 options

Journal-level data options	
Option A: Encouraging Research Data deposit and citation	Authors are encouraged to: - deposit their research data in a relevant data repository; and to - cite this dataset in their article
Option B: Encouraging Research Data deposit, citation and linking, or a Research Data Availability Statement where this is not possible. **DEFAULT**	Authors are encouraged to: - deposit their research data in a relevant data repository; - cite and link to the dataset in their article; and where this is not possible to - Make a statement explaining why research data cannot be shared.
Option C: Requiring Research Data deposit, citation and linking or a Research Data Availability Statement where this is not possible.	Authors are required to: - deposit their research data in a relevant data repository; - cite and link to the dataset in their article; or - Make a statement explaining why the research data cannot be shared.
Option D: Requiring Research Data deposit, citation and linking	Authors are required to: - deposit their research data in a relevant data repository; - cite and link to the dataset in their article
Option E: Requiring Research Data deposit, citation and linking or a Research Data Availability Statement where this is not possible. Research Data will be peer reviewed prior to publication.	Authors are required to: - deposit their research data in a relevant data repository; and to - cite and link to the dataset in their article. In addition: - Peer reviewers are asked to review the data prior to publication*

Making it easy for authors to share data



Linking research data

Link your research dataset

Please enter all details for datasets underlying your article below. To link to data or entities via accession numbers within your manuscript, please read instructions here.

Repository name

Dataverse

Dataset URL

https://dx.doi.org/10.10.1038/nphys1170

Persistent identifier

Enter the unique identifier for your dataset provided by the data repository:

DOI

10.10.1038/nphys1170

Source

Are you linking to original data reported in your research, or are you linking to reference data which is supporting your research article?

Original data Reference data

Title

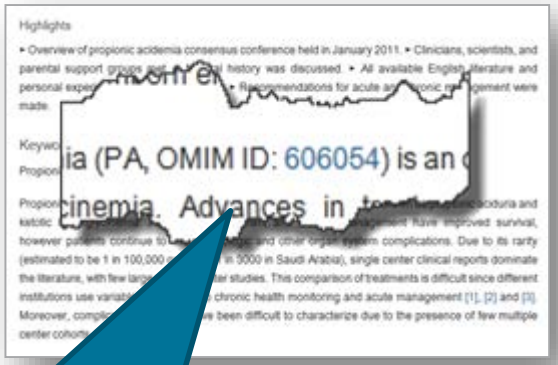
Thermodynamics

Description

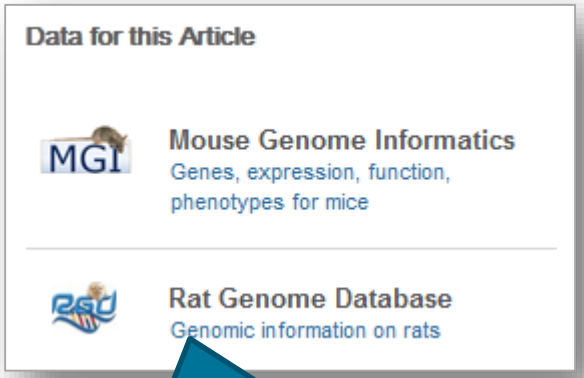
The nature of heat transfer

Linking articles to external data repositories

- Free Data-linking tool provides one click access to relevant stored data sets
- 50+ leading domain specific data repoistories
- Linking enabled by in-article accession numbers, data DOI's, or data banners



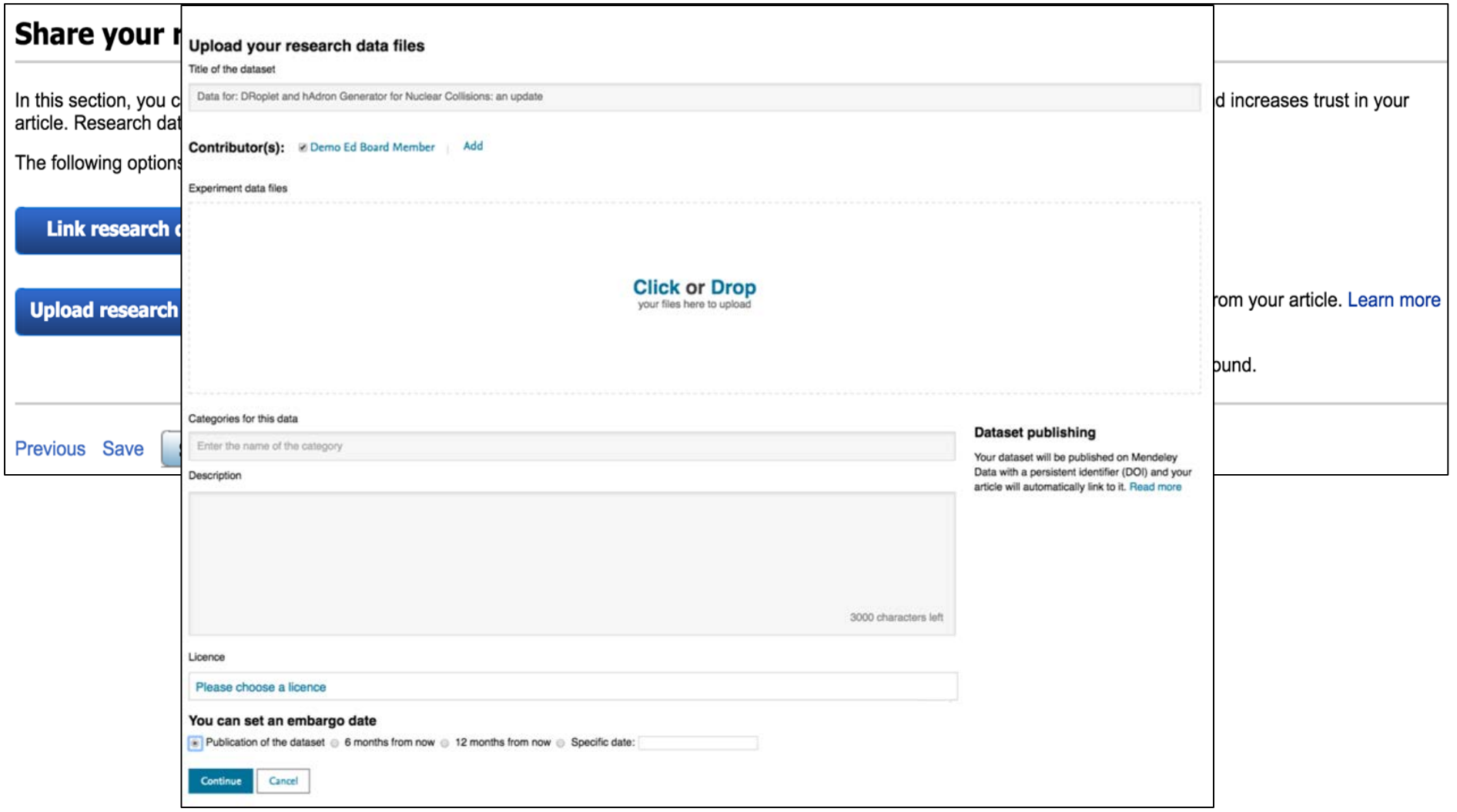
Linking through **in-article data accession numbers**



Database banners shown next to the article on ScienceDirect

See http://www.elsevier.com/databaselinking

Uploading research data



Linked to published papers – or not

Linked to Github – or not

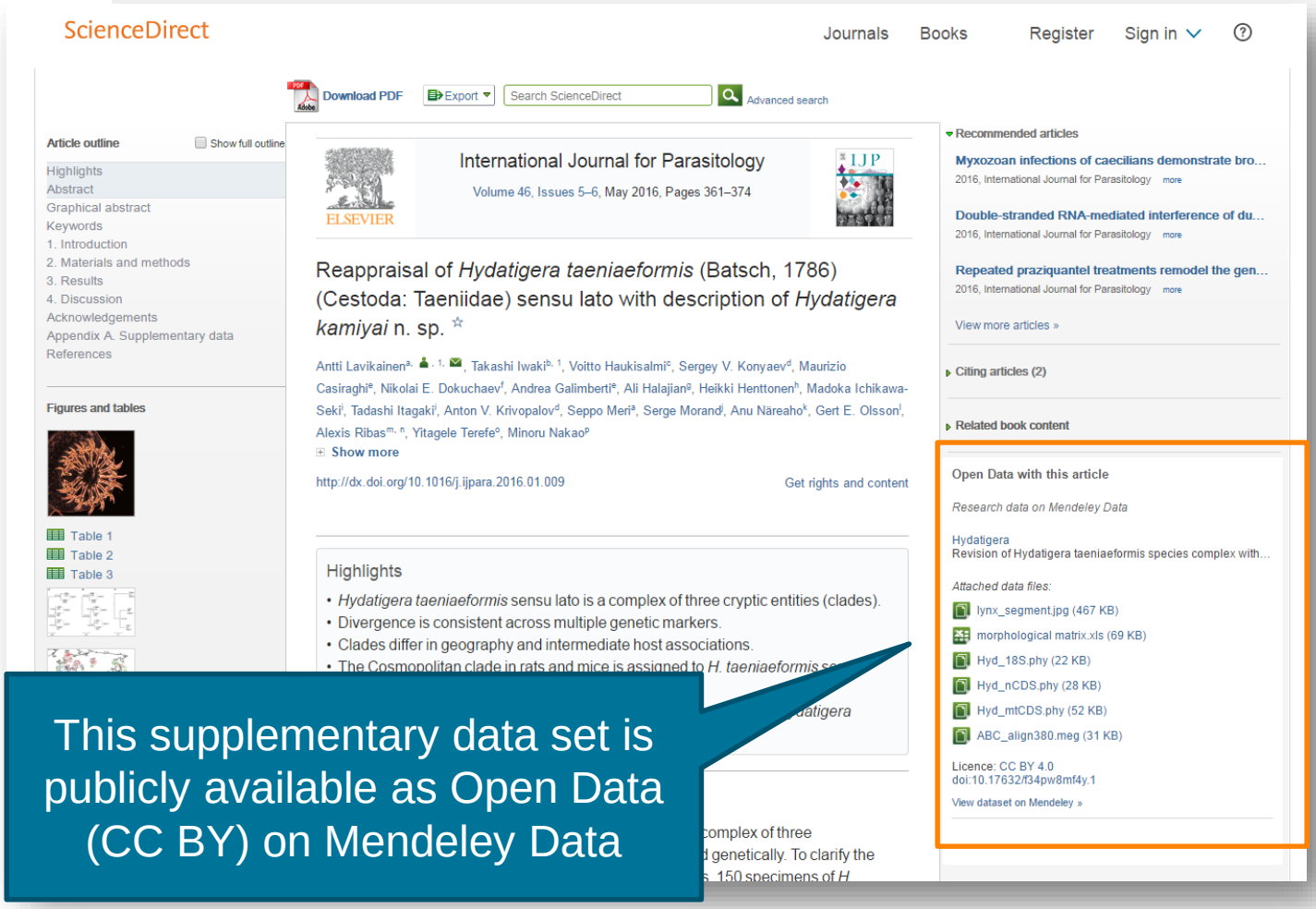
Versioning and provenance

Allowing Different Licenses

MENDELEY DATA

https://data.mendeley.com/

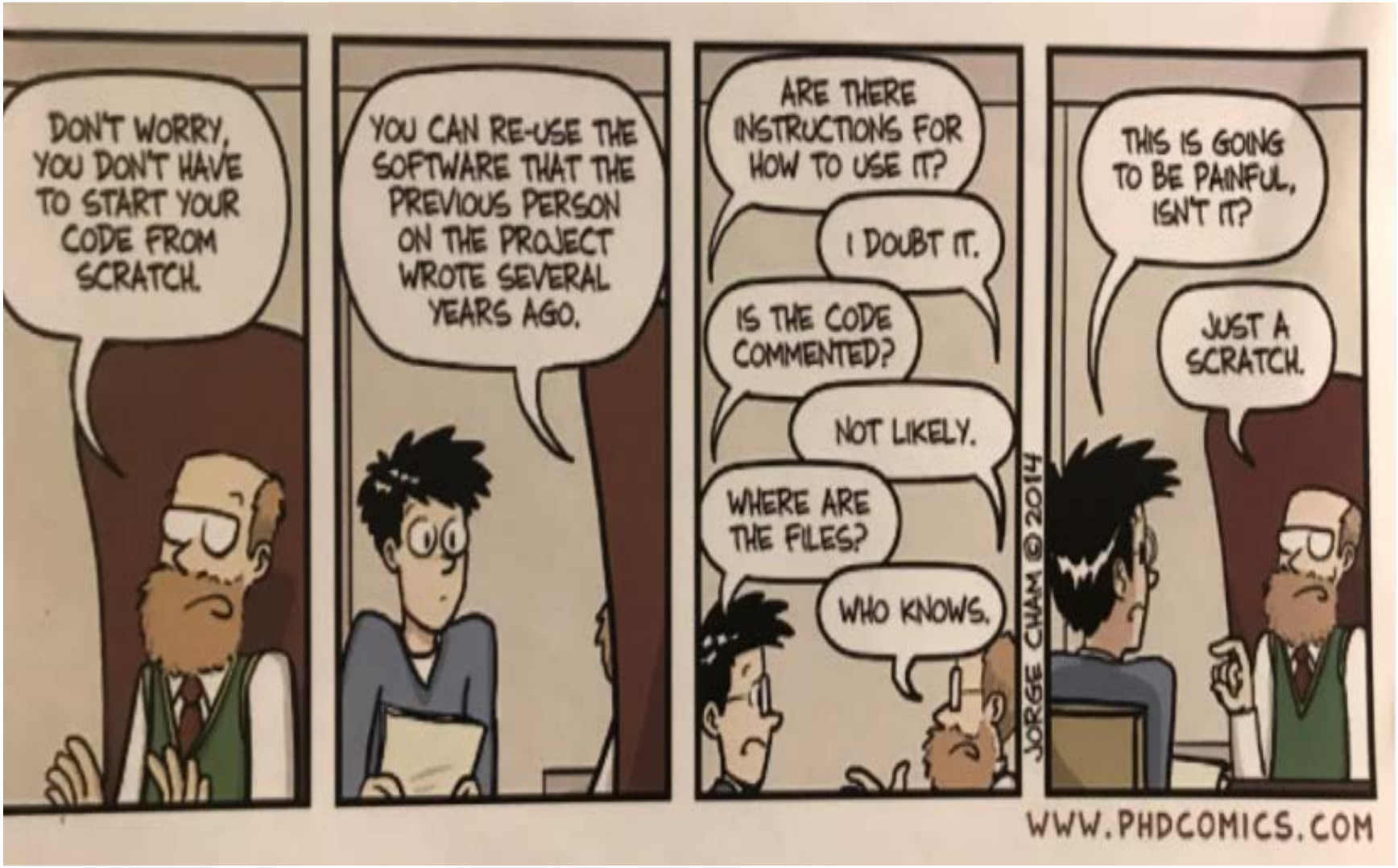
Linking open data from ScienceDirect



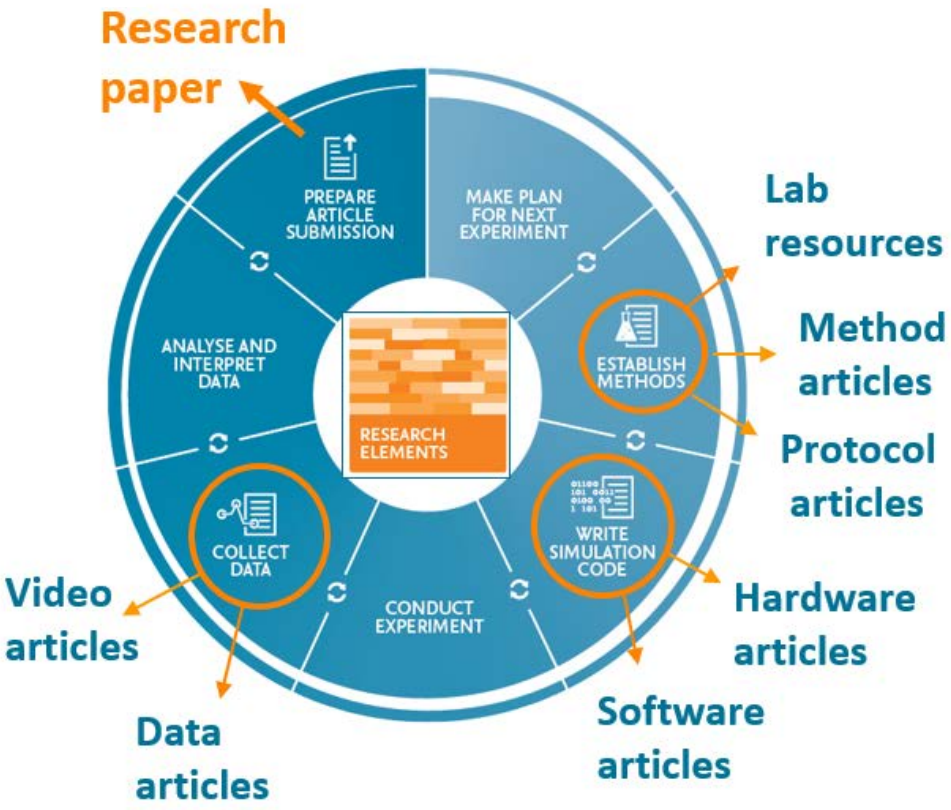
View article: http://www.sciencedirect.com/science/article/pii/S0020751916000588

Providing a statement when data cannot be shared

- No data was used for the research described in the article
- The data that has been used is confidential
- The authors do not have permission to share data
- Data will be made available on request
- The authors are unable or have chosen not to specify which data has been used
- Other (please explain in comments)


DON'T WORRY, YOU DON'T HAVE TO START YOUR CODE FROM SCRATCH.
YOU CAN RE-USE THE SOFTWARE THAT THE PREVIOUS PERSON ON THE PROJECT WROTE SEVERAL YEARS AGO.
ARE THERE INSTRUCTIONS FOR HOW TO USE IT?
I DOUBT IT.
IS THE CODE COMMENTED?
NOT LIKELY.
WHERE ARE THE FILES?
WHO KNOWS.
JORGE CHAM © 2014
THIS IS GOING TO BE PAINFUL, ISN'T IT?
JUST A SCRATCH.
WWW.PHDCOMICS.COM

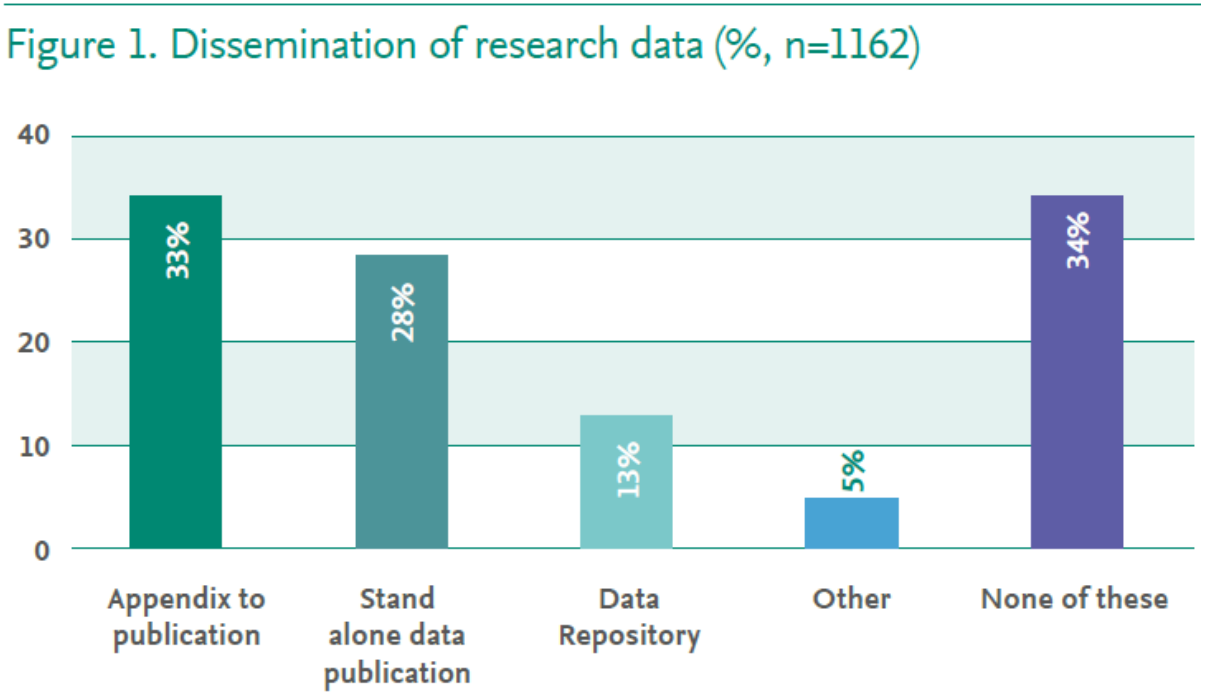
Publishing data articles



- Offer researchers an easy channel to publish their research output, **receive credit**, and **make research objects discoverable**
- Dedicated sections in regular journals
- Specialized, fast and transparent peer-review process
- Co-submission service via regular journals for data articles
- Primarily **Open Access,** APCs of ~500 USD

Because no matter what field of research you are in, DATA, METHODS and SOFTWARE always matter

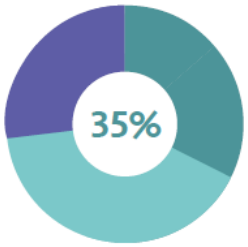
1/3 of respondents doesn't share data



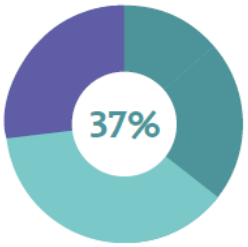
Q: Have you published the research data that you used or created as part of your project in any of the following ways?

….obstacles remain

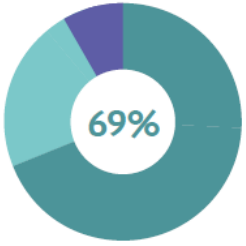
Figure 2. Attitudes towards sharing of research data (%, n=1162)



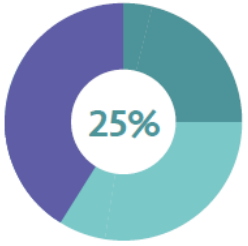
I encounter obstacles sharing my research data with others



Sharing research data is associated with credit or reward in my field



Sharing research data is important for doing research in my field



I have received sufficient training in research data sharing

Strongly agree/Agree | Disagree/Strongly disagree | Neither agree or disagree/Don't Know

Data Citation – credit for sharing

FORCE11 The Future of Research Communications and e-Scholarship

DC1 Data Citation Principles

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

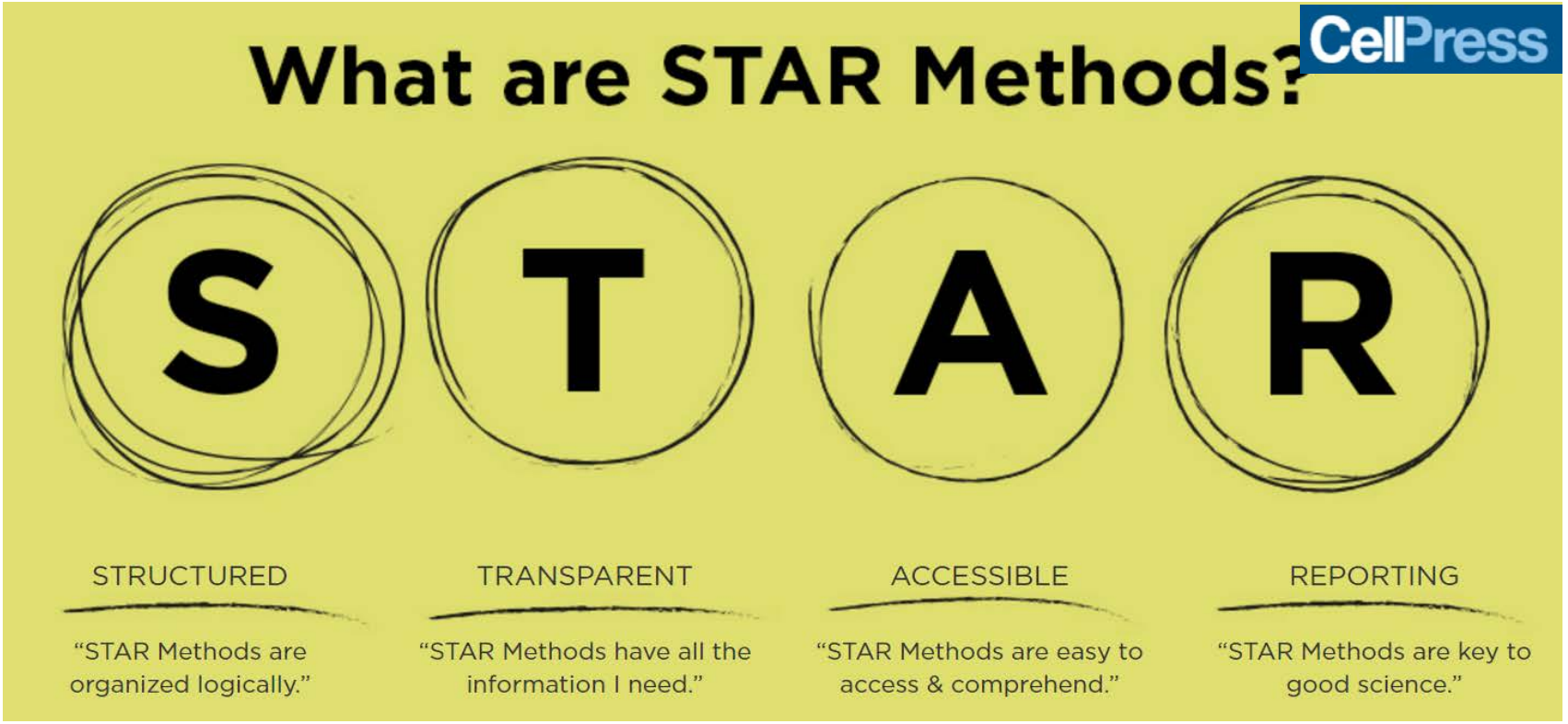
Appl Radiat Isot, 66 (2008), pp. 1745–1749

Article | PDF (272 K) | View Record in Scopus | Citing articles (27)

A Data Citation Roadmap for Scientific Publishers, https://doi.org/10.1101/100784

At manuscript submission: Methods and Materials

- **Large cause of reproducibility problems is incorrect, incomplete, or even unavailable materials and methods. We need ways to ensure that such information is properly provided by authors.**



Cell Press introduced STAR methods in 2016

- **Methods are of high value**
  - Key for evaluating merit of a manuscript
- **The truth is in the details**
  - Critical methods details are often relegated to supplement or not even there at all
- **Calls for reproducibility**
  - Without complete methods, reproducibility impossible
  - Address changing standards for methods reporting

CellPress

Main text (both print and e) includes outline M&M section with hyperlinks

New M&M section with standardized and well--organized structure

Contact info for reagent sharing and links for additional data resources

M&M features support and promote NIH Principles

Simple Word template

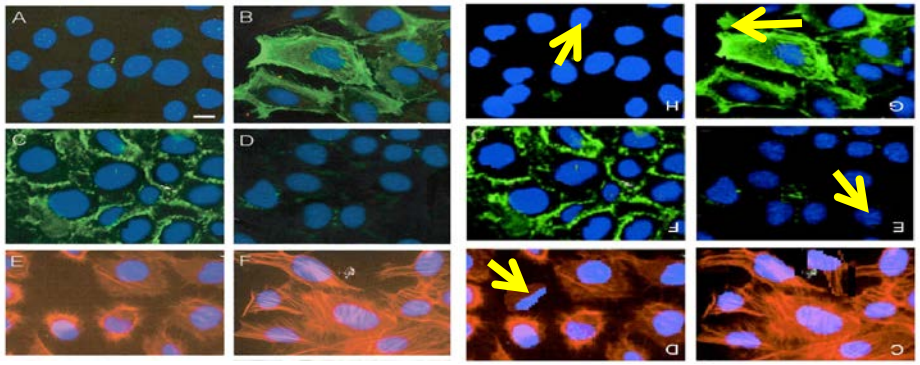
Publishing open peer review reports

- **Add optional open peer review to 1,800 journals by 2020.**
- **In January 2012, *Agricultural and Forest Metrology* began publishing selected peer review reports.**
- **A Publishing Peer Review pilot with four more journals followed, to test author, reviewer, and editor satisfaction.**
- **Reviewers were OK with their review report being published; half of them didn't require anonymity.**
- **It turns that out open peer review creates better peer review reports.**

- Reviewers who accepted the invitation:
- **95%** said publishing review reports didn't influence their **recommendation**
- **76%** said that publicly availability of their reports didn't change their **wording**
- **45%** gave consent to **reveal names**
- **36%** of those desiring anonymity said they will reveal their names **next time**
- Of reviewers who declined, **91%** said decision was **not influenced by open review**; most (68 percent) stated a lack of time as their reason for declining.

At manuscript submission: image, statistics, similarity….

- Duplication and manipulation of images (Research Integrity Solutions (RESIS):



- Validate statistics (Statcheck)

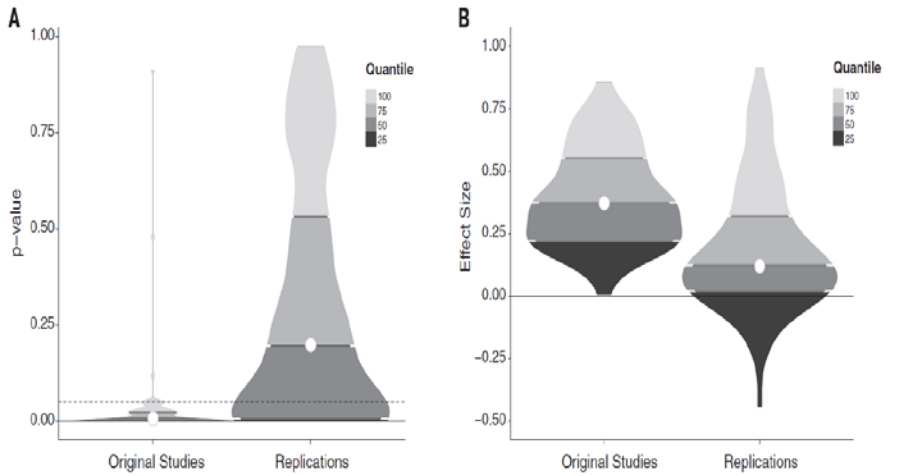


Fig. 1. Density plots of original and replication *P* values and effect sizes. (**A**) *P* values. (**B**) Effect sizes (correlation coefficients). Lowest quantiles for *P* values are not visible because they are clustered near zero.

Crossref Similarity Check

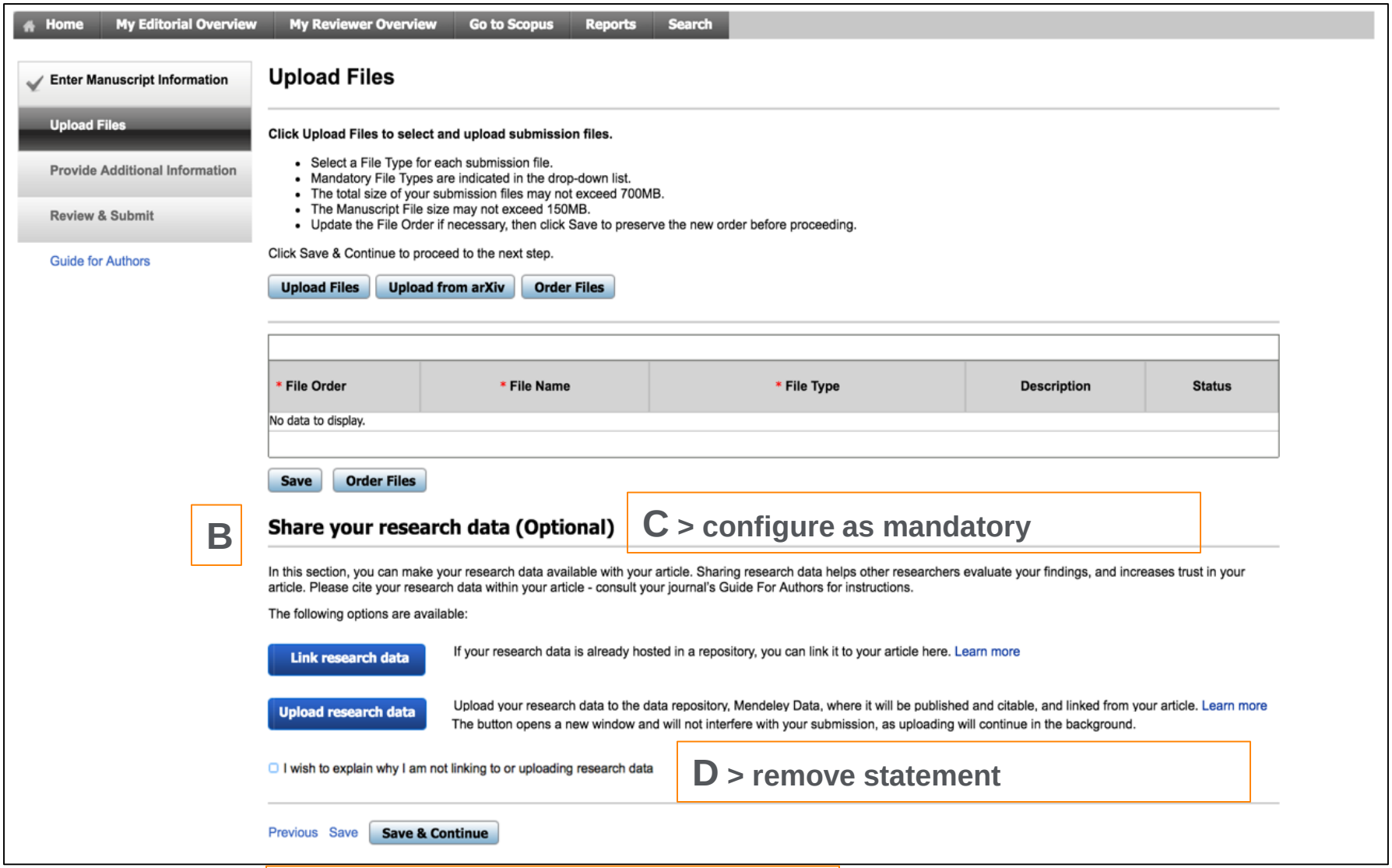
Thank you!

f.rosetta@elsevier.com

Scientific Integrity – is there a problem? There are different opinions out there …

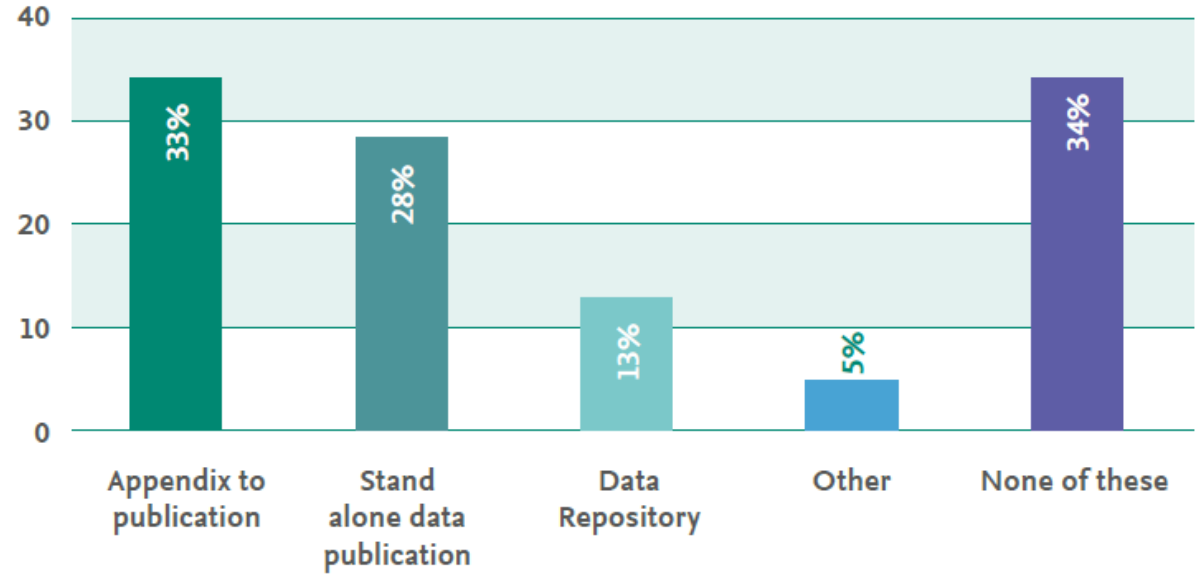
- 2% scientists admitted FFP (fabrication, falsification, and plagiarism) at least once; and 14% saw colleagues falsify outcomes. [Fanelli, 2009]
- 72% scientists saw questionable research practices (QRP) with colleagues. [Fanelli, 2009]
- Ioannidis “charges that as much as 90 percent of the published medical information that doctors rely on is flawed”. [The Atlantic, Nov 2010]
- In 2012, PubMed had 25 million articles and only 2047 retractions (0.01%). [Steen et al, 2013]

Implementing the data guidelines



1/3 of respondents doesn't share data

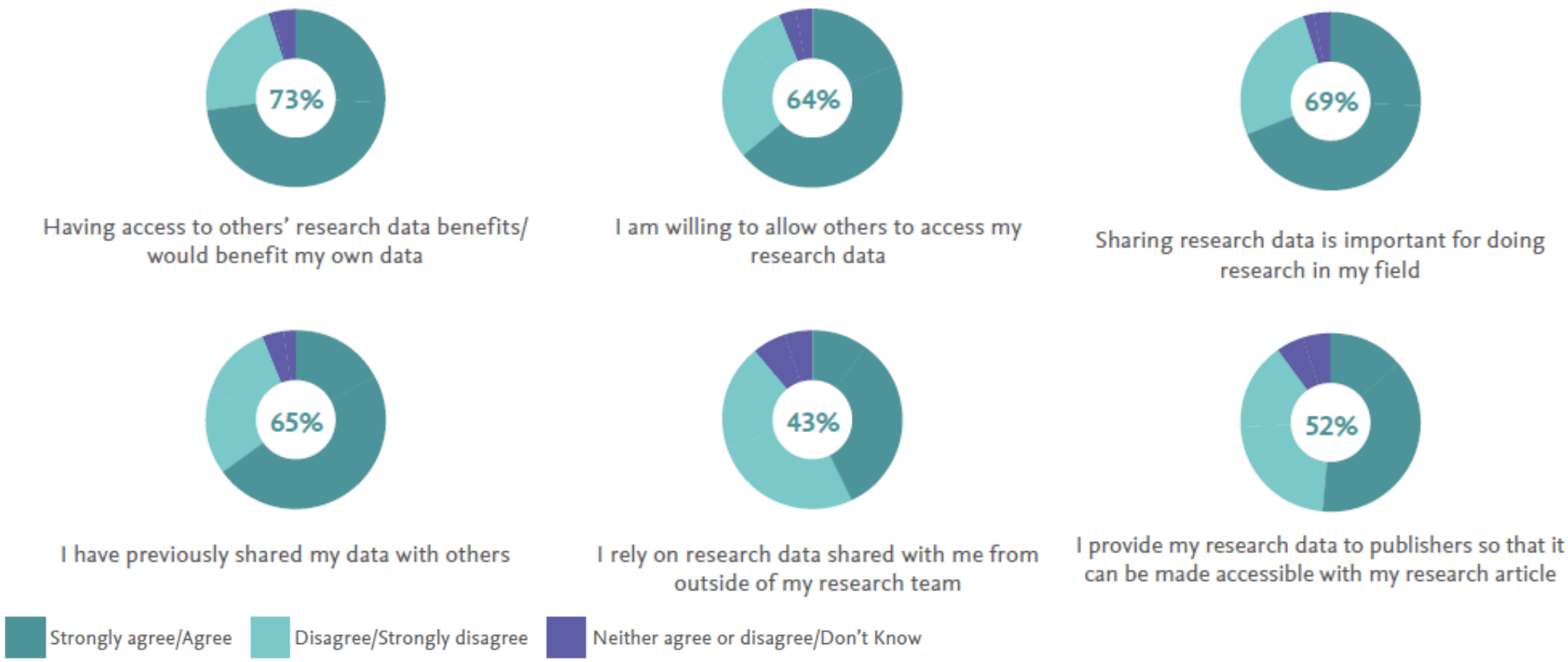
Figure 1. Dissemination of research data (%, n=1162)



Q: Have you published the research data that you used or created as part of your project in any of the following ways?

Even though advantages of sharing data are clear…

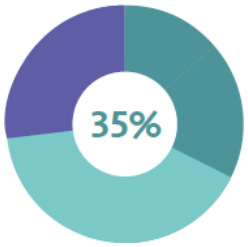
Figure 2. Attitudes towards sharing of research data (%, n=1162)



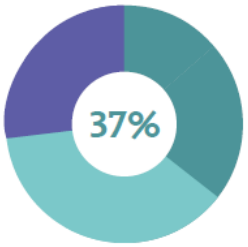
Q: To better understand your attitudes towards research data access, please think about the research data that typically is not published (e.g. not summary charts, tables or images), and indicate how much you agree or disagree with the following statements.

….obstacles remain

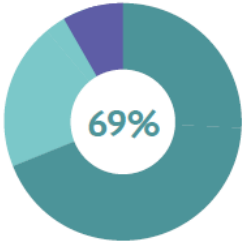
Figure 2. Attitudes towards sharing of research data (%, n=1162)



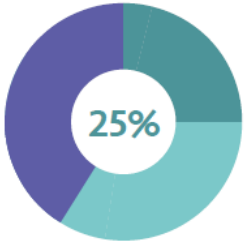
I encounter obstacles sharing my research data with others



Sharing research data is associated with credit or reward in my field



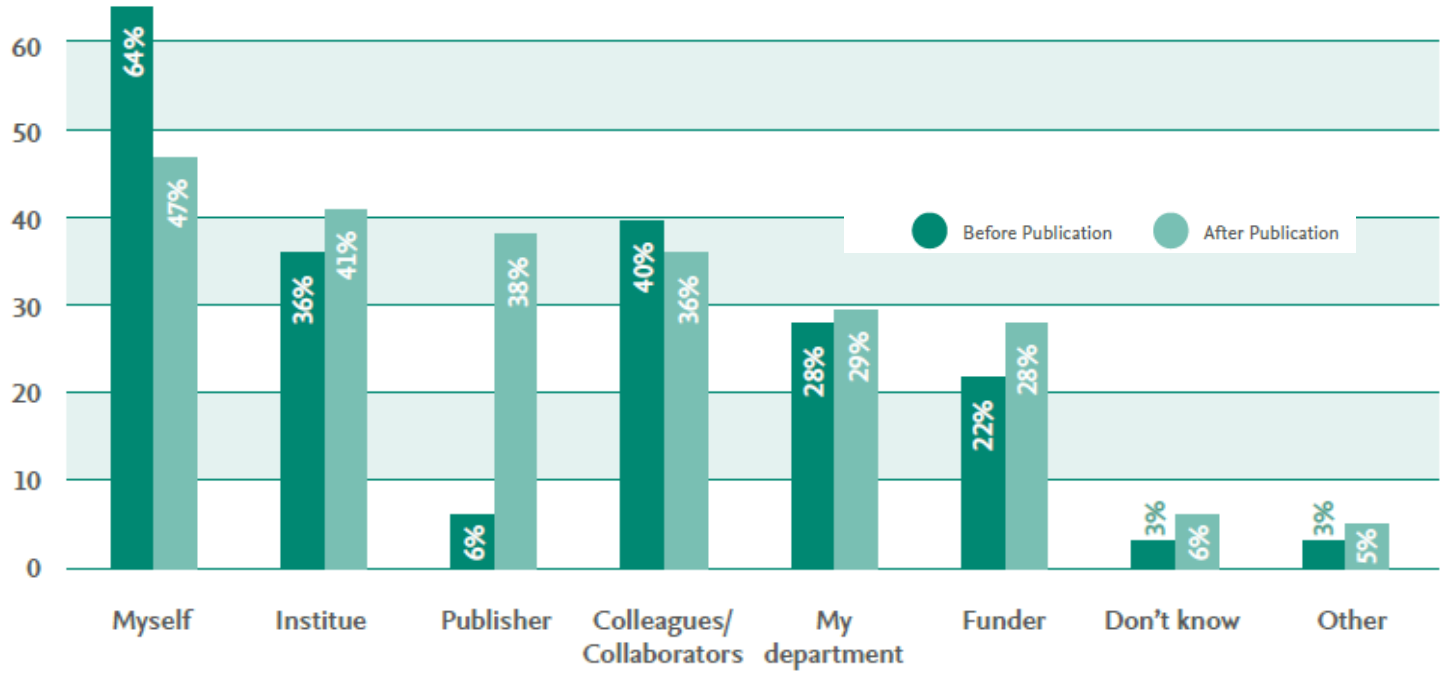
Sharing research data is important for doing research in my field



I have received sufficient training in research data sharing

Little clarity on who own rights on data

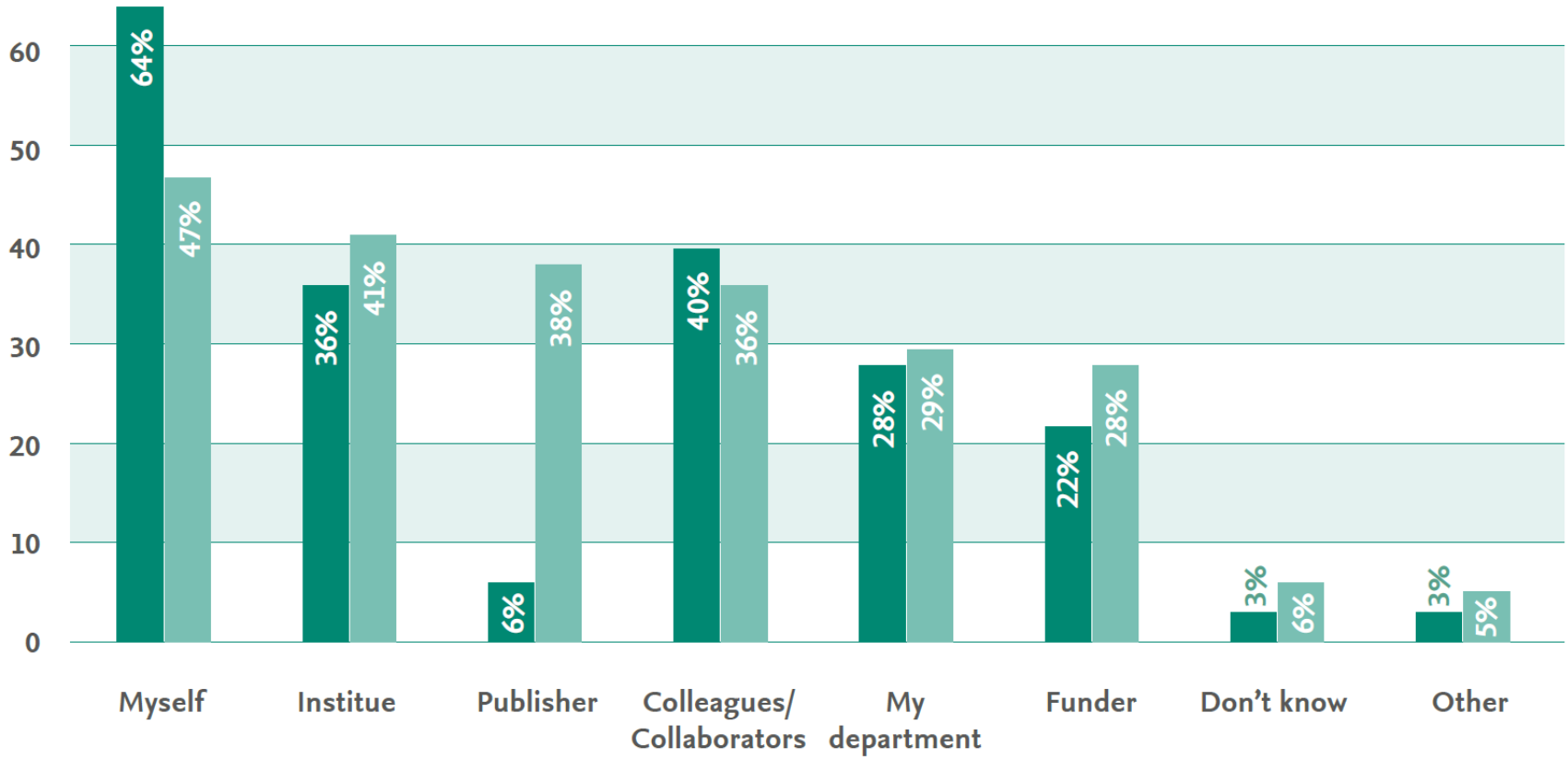
Figure 3. Research data ownership before and after publication (%, n=1162)



Q: Who do you believe 'owns' the research data that you have made or will make available to others as part of your last research project?

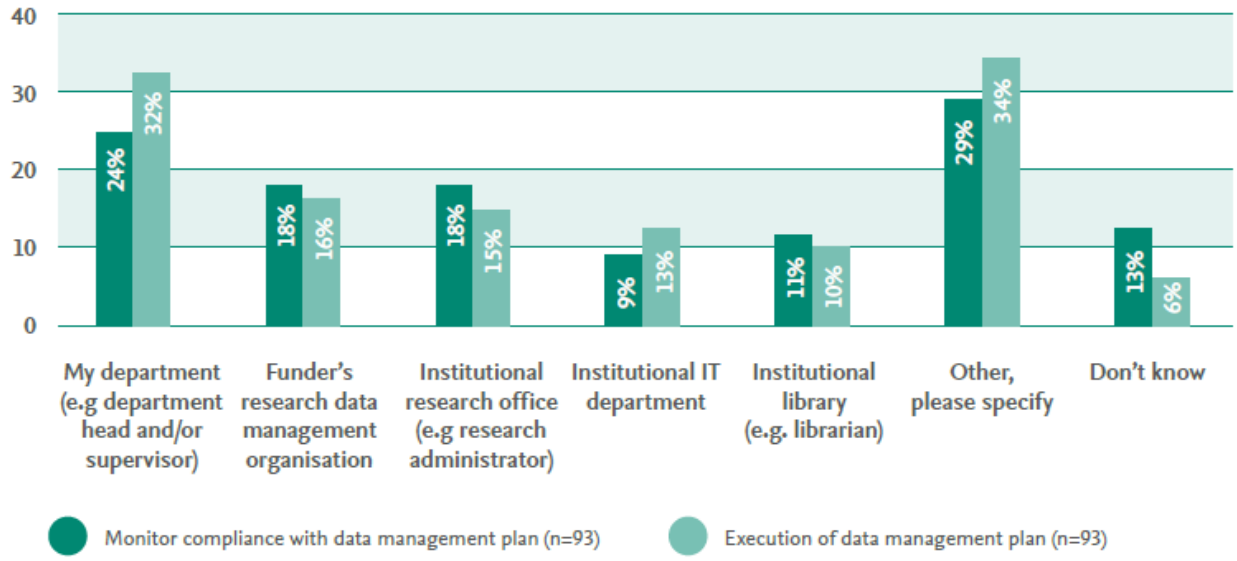
And who takes that decision?

Figure 3. Research data ownership before and after publication (%, n=1162)



Who is responsible for acting on data management plans?

Figure 4. Execution and monitoring of research data management (%)



Q: *[Respondents indicated they are mandated to archive your research data and are provided with a research data management plan to follow.]* Who is responsible for the execution this research data management plan? Who is responsible for monitoring compliance this research data management

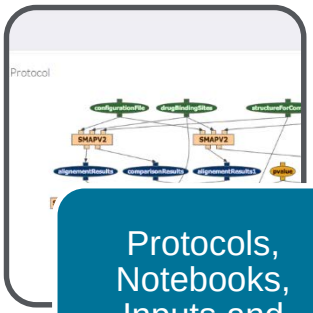
Last but not least: Reproducibility

- **The idea that replication studies are only valuable if the results disagree with the original research is a misconception, as is the idea that editors don't want to publish replication studies.**
- **Efforts on Reproducibility can be found in most of the Elsevier initiatives listed above: all data efforts, STAR methods, image manipulation detection, etc.**
- **Plus:**
  - Already in 2012, Elsevier's Dr. William Gunn co-founded The Reproducibility Initiative with Dr. Elizabeth Iorns (ScienceExchange).
  - We have developed a new article type especially for replication studies.
  - We are issuing several calls for papers to encourage submissions.

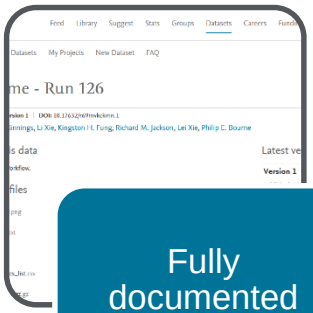
Summary

- **Yes, there are developments to be concerned about:**
  - There seems to be a light increase in retractions.
  - Another concern is the degree of reproducibility.
- **Lack of research integrity has different causes, like:**
  - Plagiarism and image issues, lower frequency, very visible.
  - Design and statistical errors, higher frequency, less visible, but potentially more harmful due to its higher frequency.
- **Core responsibility for Research Integrity is with researcher.**
- **For publishers, together with funders and institutions, there is a lot that can be done to safeguard the integrity of what we publish – and it is certain that safeguarding is needed.**
- **Where can institutions and publishers help each other?**
- **The actions should be a mix of fighting the lower-frequency high-visibility violations, and of battling the higher-frequency lower-visibility violations, which might even be more harmful.**

The challenge: use Mendeley Data Platform to record and publish experiment



Protocols, Notebooks, Inputs and Outputs captured in Notebook



Fully documented datasets shared in Mendeley Data

Dataset indexed by DataSearch.com for better discoverability

Methods paper published on MethodsX, with linked datasets and protocol

https://data.mendeley.com/datasets/r69mvkckmn/1

Elsevier Cortex was first journal with pre-registration

- Registered Reports in Cortex.
- Authors pre-register hypothesis & methods prior to doing research.

- To reduce bias against negative results, improve reproducibility and embed journals earlier in the research process.
- Press coverage in The Guardian and Elsevier Connect, interest from EC.
- Submissions do come in, though more community support needed.