Why is getting credit for your research data so hard?

2019 Research Data Management Symposium

Wouter Haak

VP Research Data Management Elsevier

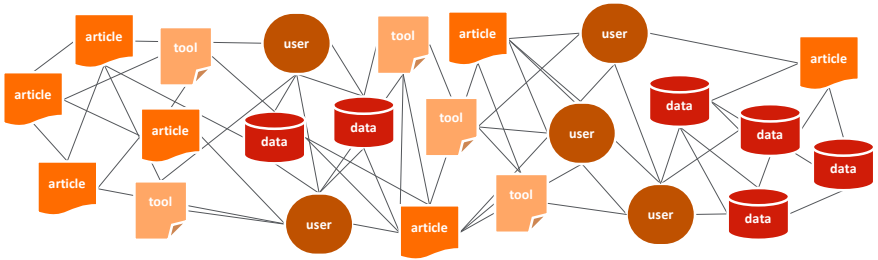
https://data.mendeley.com/

Why **publishers** care about **open science**:

Today:
linear supply chains

Linear supply chains are evolving into complex, dynamic and connected value webs

The future:
networked open science



Model: Castle

- Goal: selling content
- Metrics: number of units sold
- Strategy: optimize content delivery to users

Win by reputation

Model: Marketplace

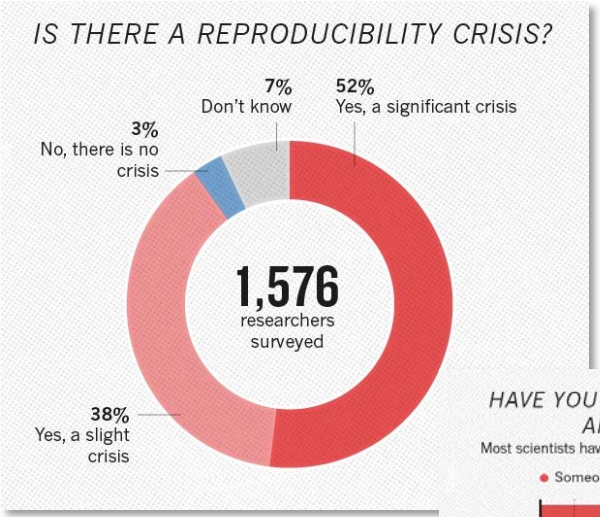
- Goal: grow number of interactions
- Metrics: number of interactions between users
- Strategy: optimize number of network interactions

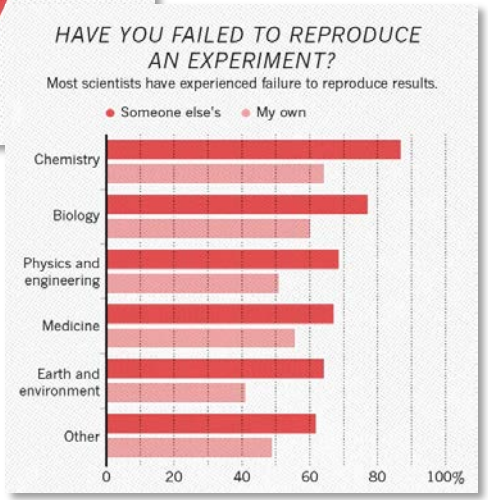
Win by trust

The reproducibility crisis

More than 50% of researchers surveyed **failed to reproduce** their own experiments

Research data doesn't just need to be available, it needs to be **comprehensible, available and trustworthy**

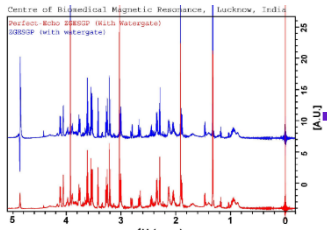




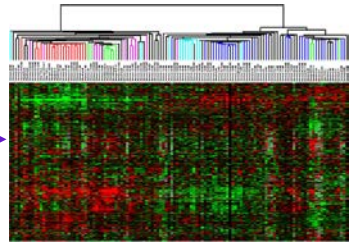
Source: https://www.nature.com/news/1-500-scientists-lift-the-lid-on-reproducibility-1.19970

When talking about data, we talk about…

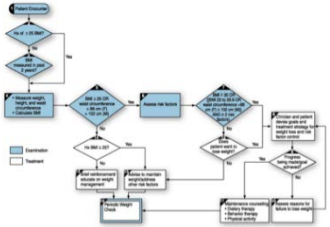
The information underpinning articles offers value to other researchers – with many now arguing that research data should be considered a "first class citizen" of research output, alongside literature publications.



Raw data

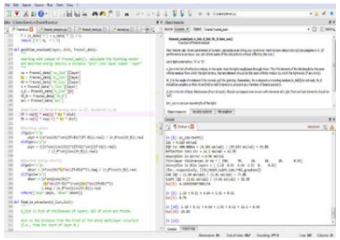


Processed data



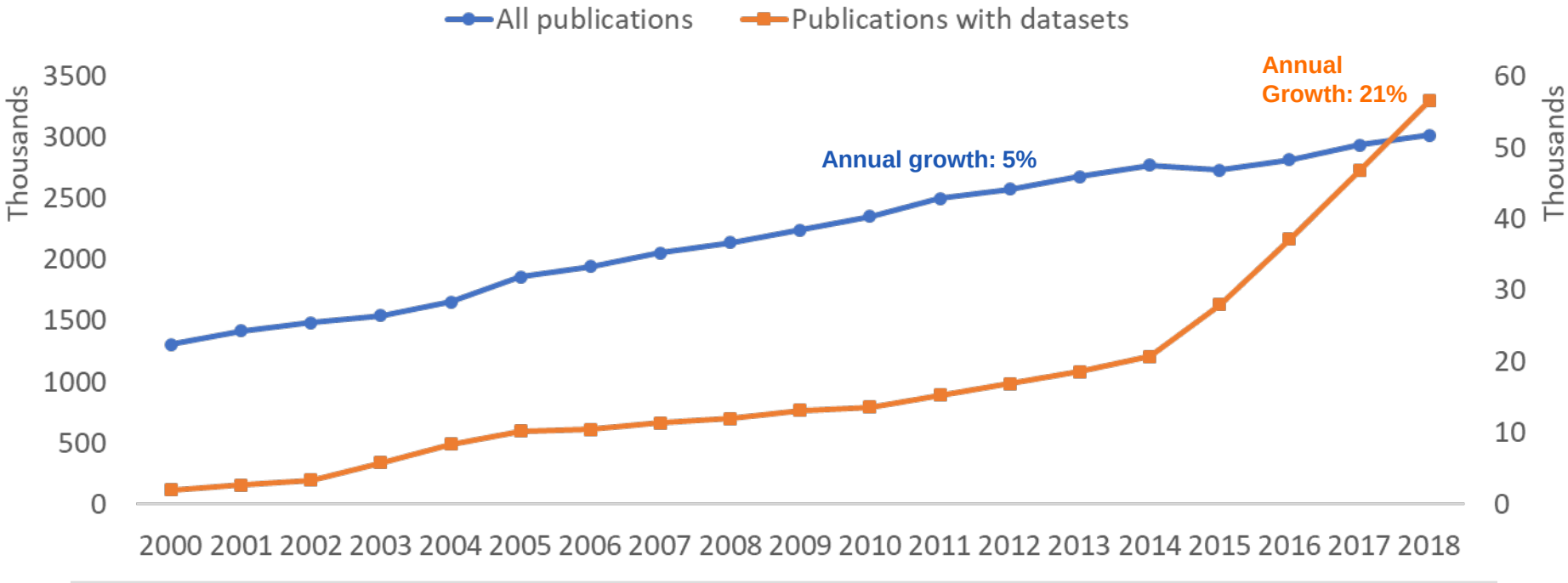
Protocols, methods, workflows

Machine & environment settings



Scripts, analyses & algorithms

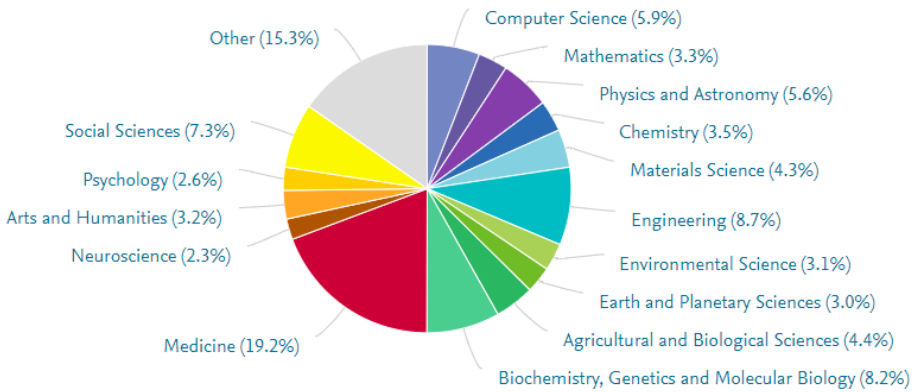
Research Data Management adoption is growing very fast worldwide



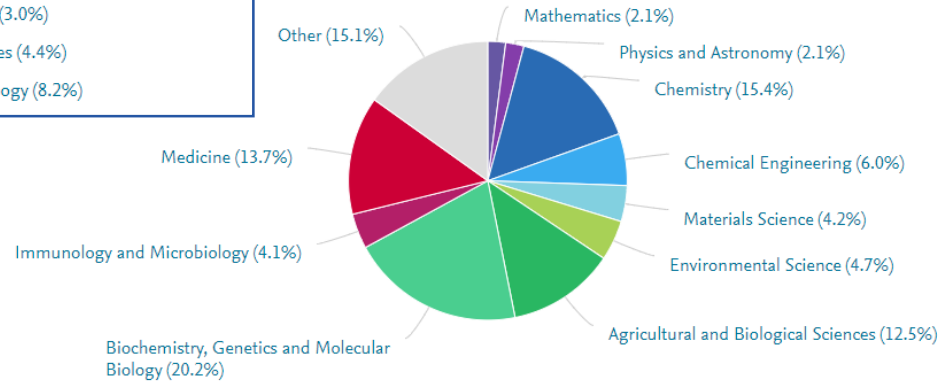
Source: Mendeley Data Monitor analysis of Scopus, Scholix, SciVal, 5 year data
2014-2018 extracted on August, 2019 – CAGR = Compound Annual Growth Rate

US: analysed 2014-2018 research articles across disciplines

In total 3,4 mln articles analysed:

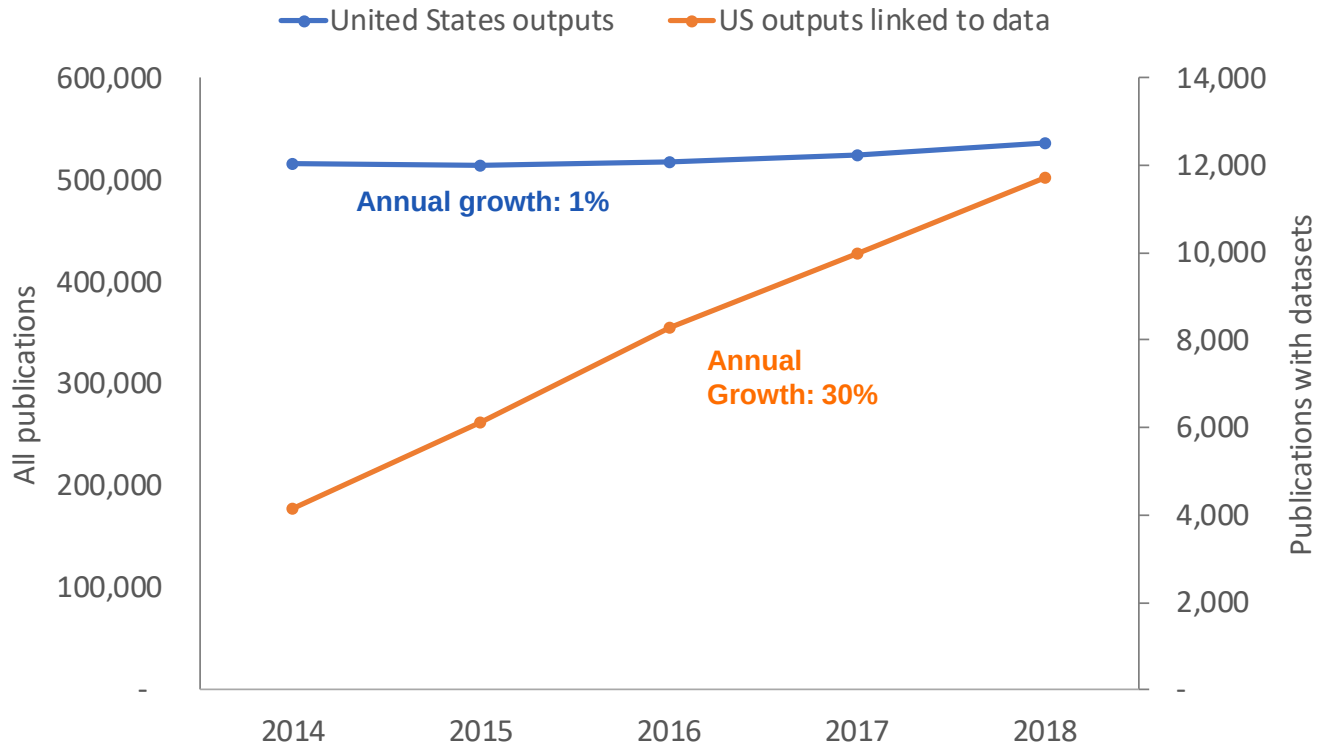


41,797 articles with associated datasets



Source: Mendeley Data Monitor analysis of Scopus, Scholix, SciVal, 5 year data 2014-2018 extracted on August, 2019 – CAGR = Compound Annual Growth Rate

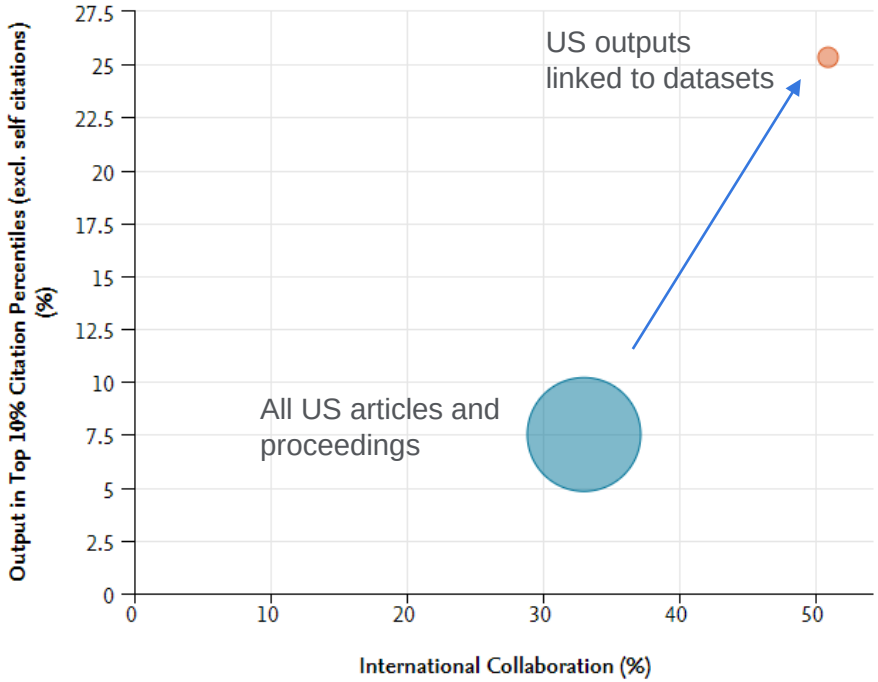
RDM adoption also growing fast in US



Source: Data Monitor analysis of Scopus, Scholix, SciVal, 5 year data 2014-2018 extracted on August, 2019 – Growth = CAGR = Compound Annual Growth Rate

Impact of sharing data in US (1)

Top cited



International collaboration

- Higher citations
- More collaborations

Source: Mendeley Data Monitor analysis of Scopus, Scholix, SciVal, 5 year data 2014-2018 extracted on August, 2019 – CAGR = Compound Annual Growth Rate

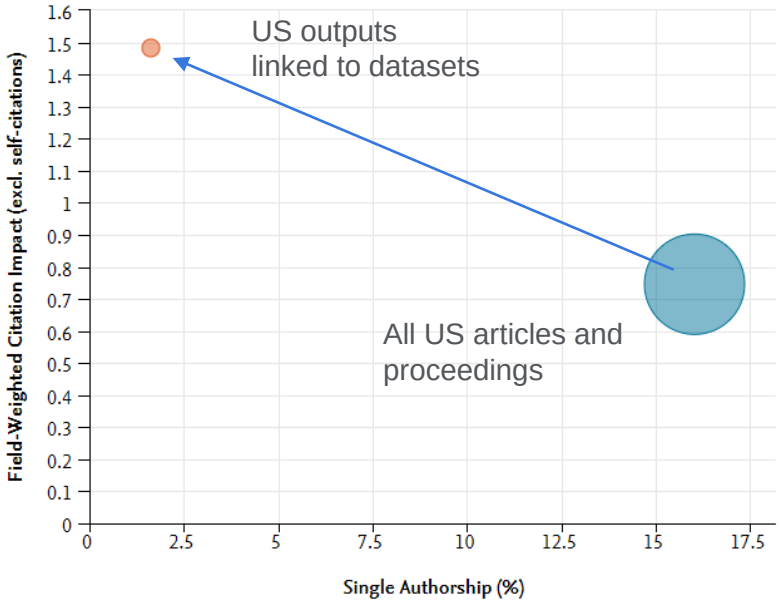
y-axis: Output in Top 10% Citation Percentiles (excl. self citations)

x-axis: International Collaboration

Bubble size: Scholarly Output

Impact of sharing data in US (2)

FWCI



Single authorship

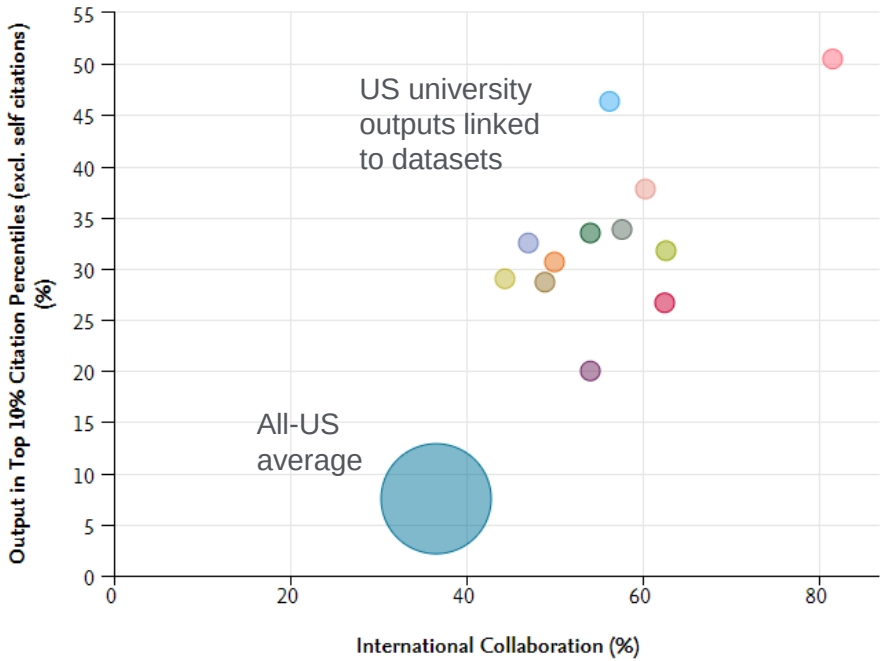
- Higher citation impact (FWCI)
- More collaborations (less single authors)

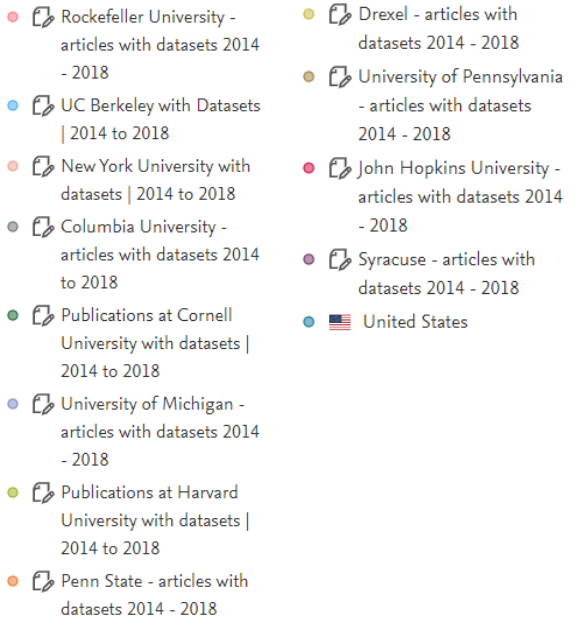
Source: Mendeley Data Monitor analysis of Scopus, Scholix, SciVal, 5 year data 2014-2018 extracted on August, 2019 – CAGR = Compound Annual Growth Rate

y-axis: Field-Weighted Citation Impact (excl. self-citations)
x-axis: Single Authorship
Bubble size: Scholarly Output

Data sharing helps all US universities - some examples

Top cited





International collaboration

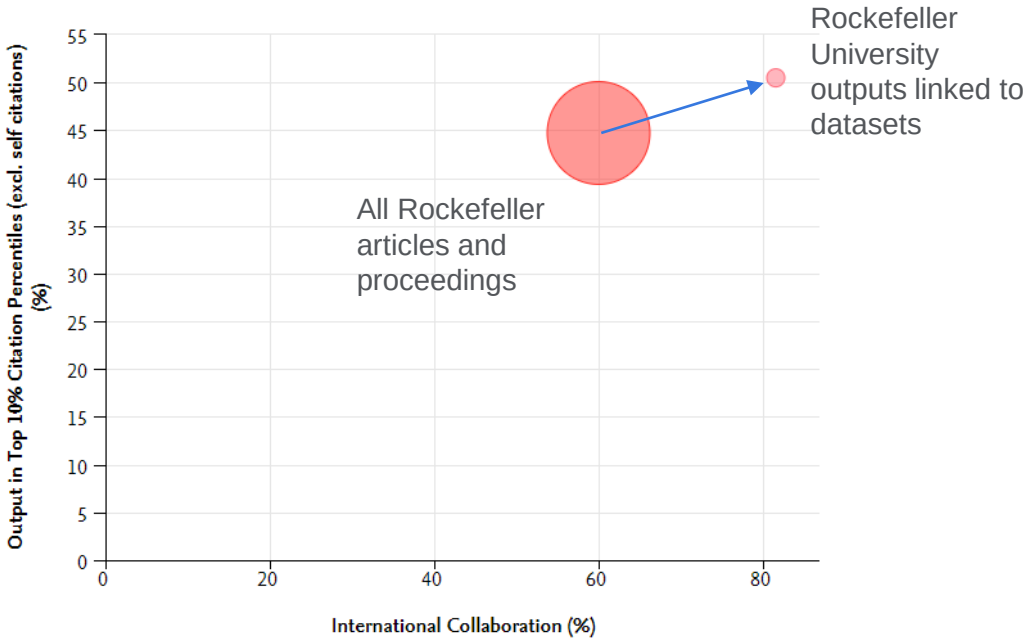
Source: Data Monitor analysis of Scopus, Scholix, SciVal, 5 year data 2014-2018 extracted on August, 2019 – CAGR = Compound Annual Growth Rate

y-axis: Output in Top 10% Citation Percentiles (excl. self citations)

x-axis: International Collaboration

Bubble size: Scholarly Output

Datasharing at Rockefeller University: impact!



Source: Data Monitor analysis of Scopus, Scholix, SciVal, 5 year data 2014-2018 extracted on August, 2019 – CAGR = Compound Annual Growth Rate

y-axis: Output in Top 10% Citation Percentiles (excl. self citations)

x-axis: International Collaboration

Bubble size: Scholarly Output

Sharing data works:
25% higher citation impact

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.

* g.colavizza@uva.nl

https://arxiv.org/pdf/1907.02565.pdf

Abstract

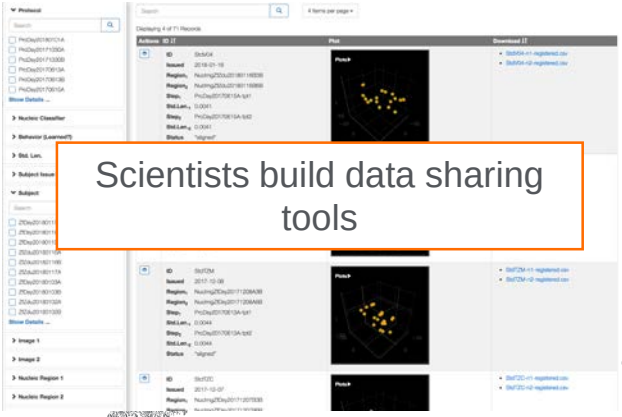
Efforts to make research results open and reproducible are increasingly reflected by journal policies encouraging or mandating authors to provide data availability statements. As a consequence of this, there has been a strong uptake of data availability statements in recent literature. Nevertheless, 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 them. We consider 531,889 journal articles published by PLOS and BMC which are part of the PubMed Open Access collection, categorize their data availability statements according to their content and analyze the citation advantage of different statement categories via regression. We find that, following mandated publisher policies, data availability statements have become common by now, yet statements containing a link to a repository are still just a fraction of the total. We also find that articles with these statements, in particular, can have up to 25.36% higher citation impact on average: an encouraging result for all publishers and authors who make the effort of sharing their data. All our data and code are made available in order to reproduce and extend our results.

Some examples of *Open Data and Open Science*:

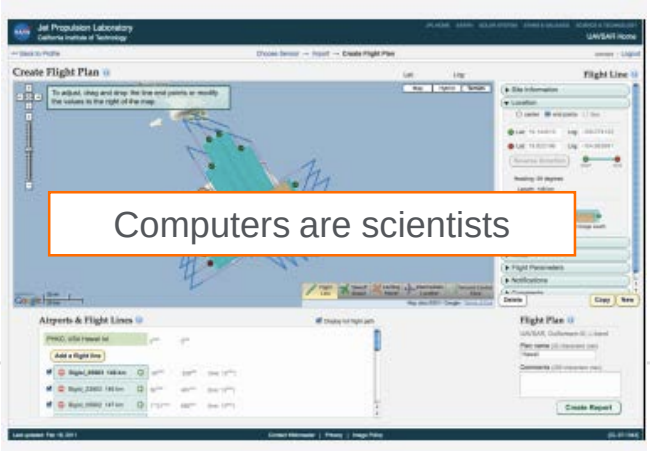
Carl Kesselman builds tools to enable neuroscientists to store and share their data in a better way

Viktor Pankratius builds software programs that generate hypotheses about volcano eruptions: the software can steer drones to collect data.

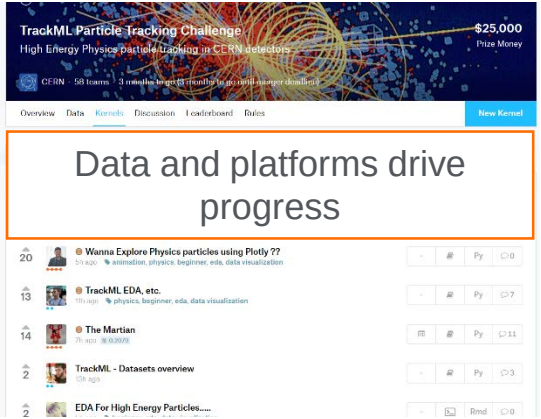
Lena Deus solves scientific problems through Kaggle: the system awards her points for scoring highest on Machine Learning tasks.



Scientists build data sharing tools



Computers are scientists



Data and platforms drive progress

End-to-end RDM

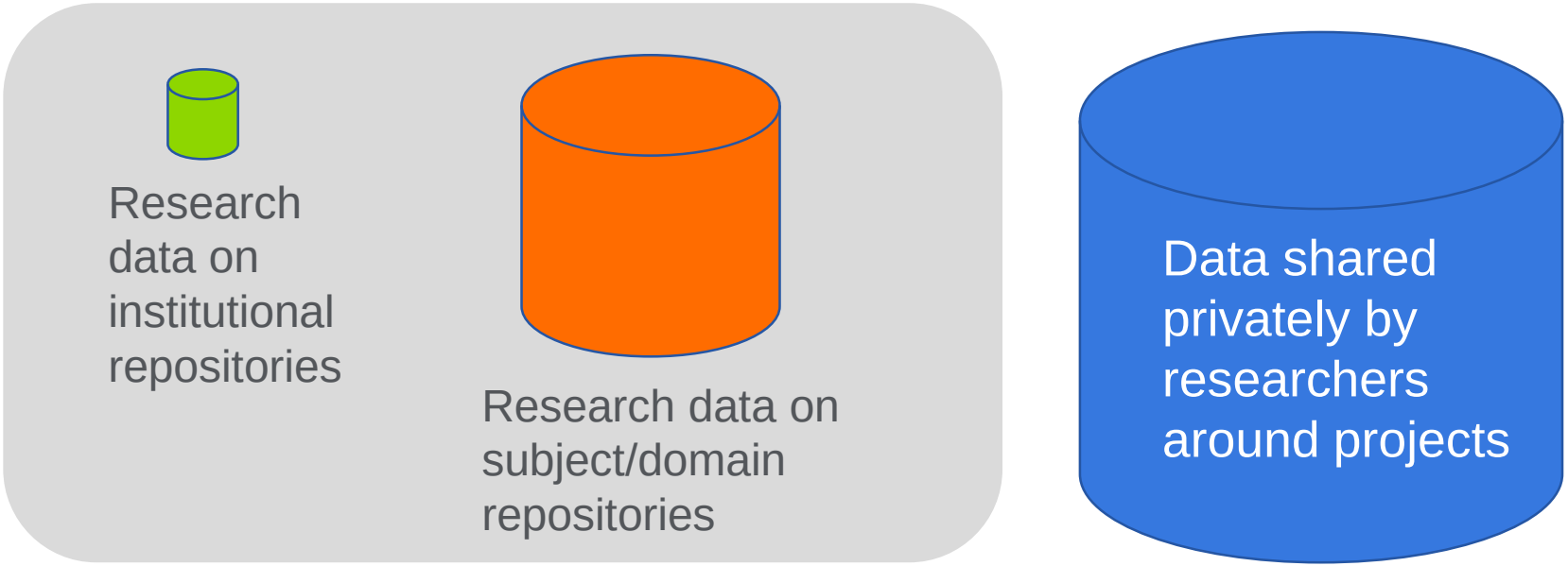
Organizing for RDM: Pitfall 1 = Admin

- 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/

Organizing for RDM: pitfall 2 = Assume all research data is at your institution

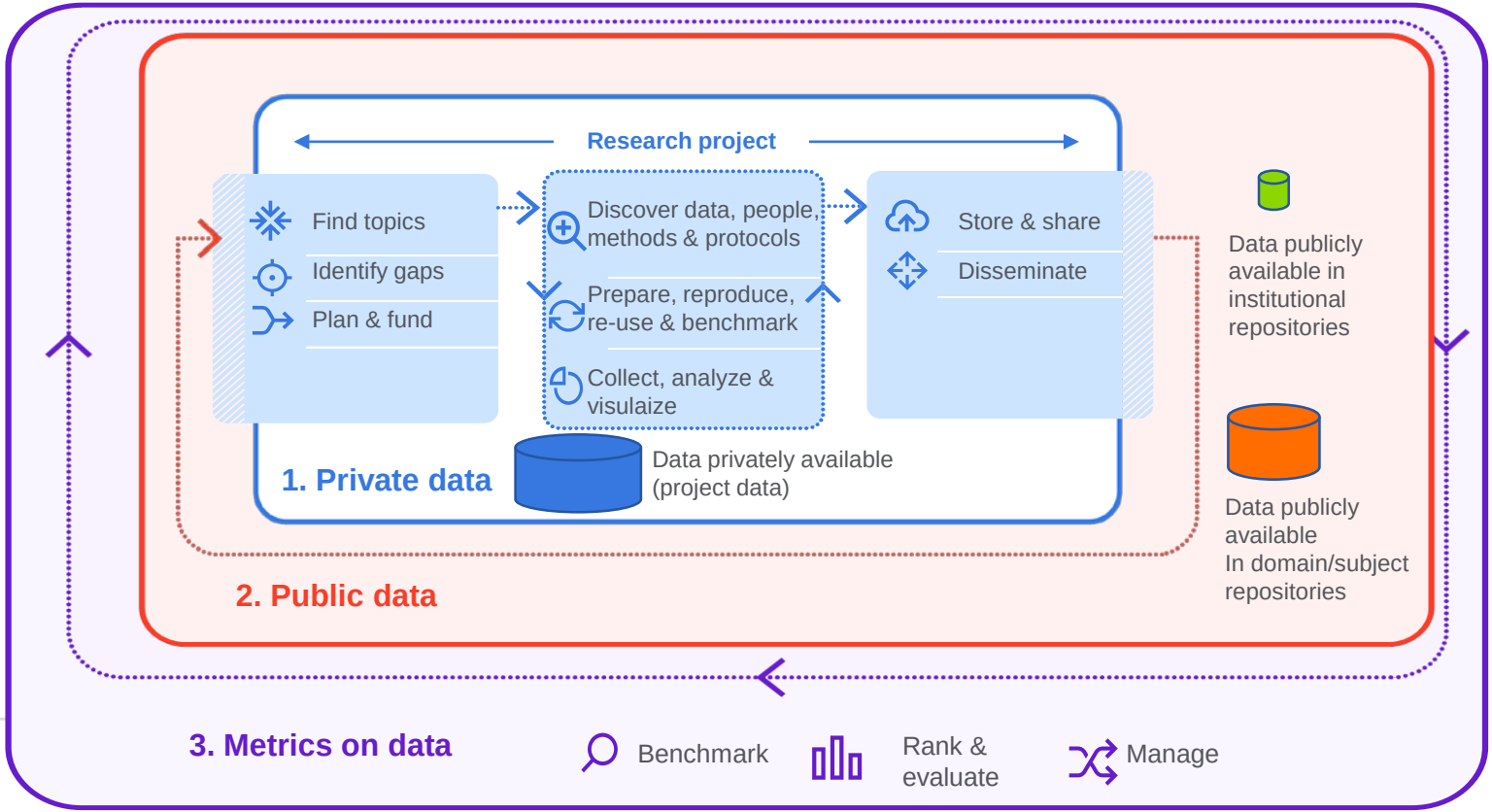


Organizing for RDM: pitfall 3 = Assume private data is reusable in the future

- Is research data a strategic asset for the future of your institution?
- Are your researchers preserving data for future reuse?
- What happens when a researcher leaves?
- Do you have an overview of data at your institution?

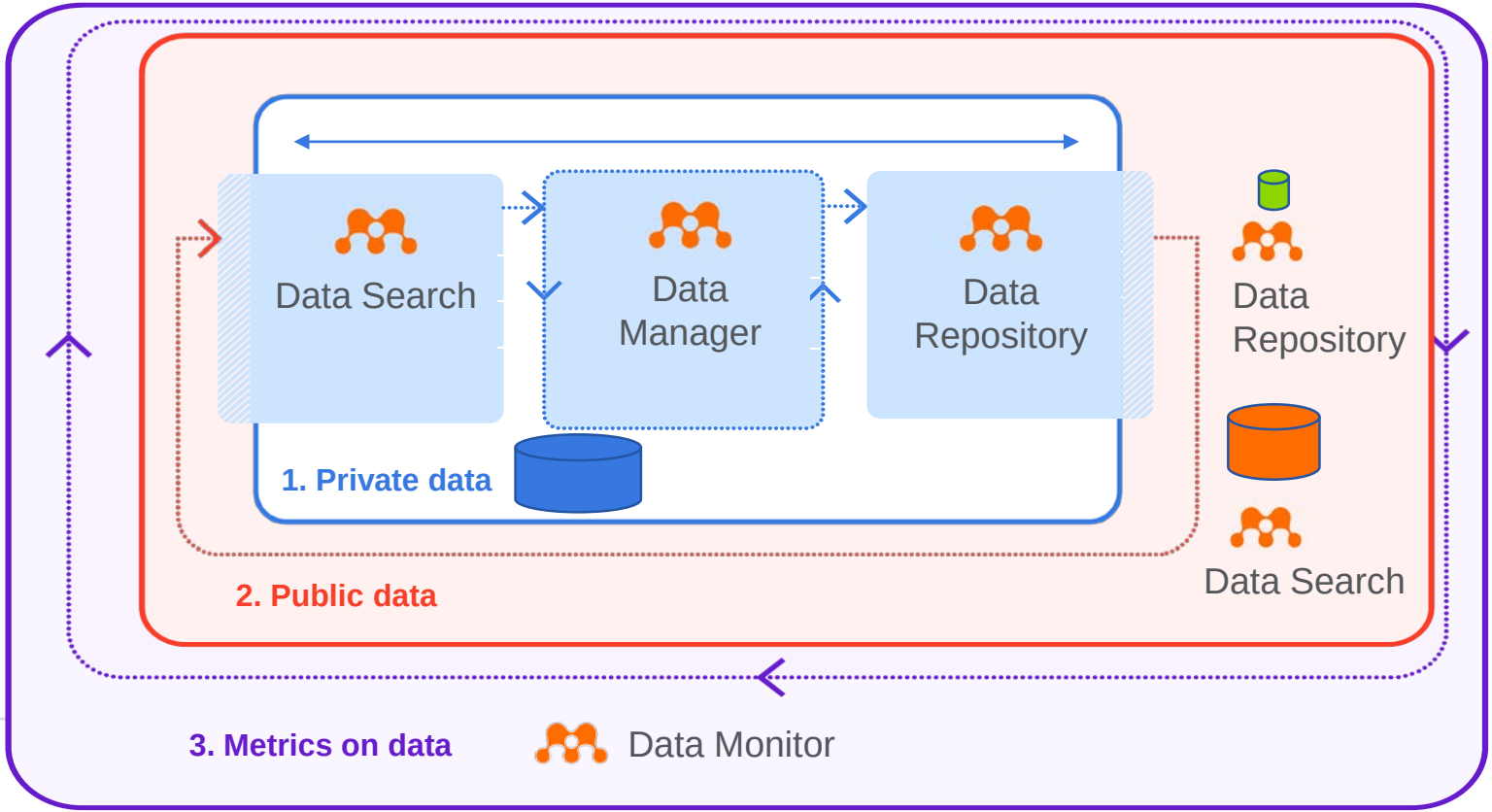
Private Research data

RDM: need to support three data life-cycles



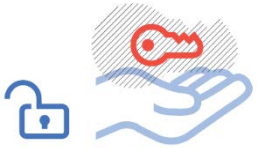
ELSEVIER

Supporting three data life-cycles

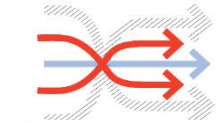


Five Facts about Elsevier and Research Data

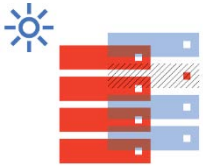
Fact #1 Elsevier's Mendeley Data supports the entire lifecycle of research data
The **4 modules** that make up Mendeley Data are specifically designed to utilize data to its fullest potential, simplifying and enhancing current way of working.



Fact #2 Researchers and institutions own and control all the data
Mendeley Data allows researchers to keep data private, or publish it under one of **16 open data licenses**, so they stay in full control



Fact #3 Mendeley Data is an open system
It is a **flexible platform** — modules are designed to be used together, standalone, or combined with other Elsevier and non-Elsevier solutions



Fact #4 Mendeley Data can increase the exposure and impact of research
Mendeley Data Search indexes over **10 million datasets** from more than **35 repositories**

Fact #5 Elsevier is an active participant in the open data community
Elsevier partners with the open data community, and is currently working on more than **20 projects globally**

ELSEVIER

Why is getting credit for your research data so hard?

Perhaps it is less hard than you think: good things are already happening

Thank you

w.haak@Elsevier.com