



# Focus on Open Science- Where do we stand: Open Data/FAIR Data

Nov 2019  
Regional Manager Adam Goh



# Why is Research Data Management (RDM) “hot” now?

## 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

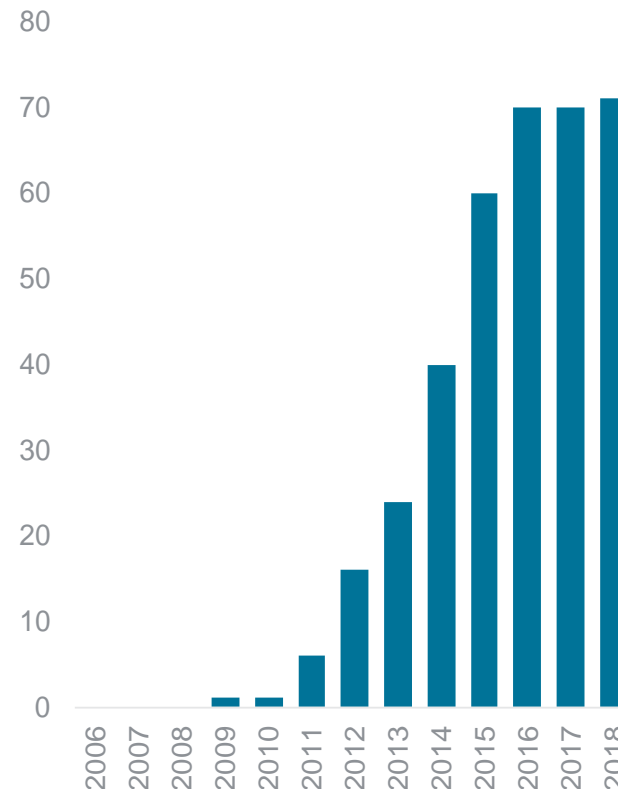
### 2) Funding bodies & institutions:

- 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

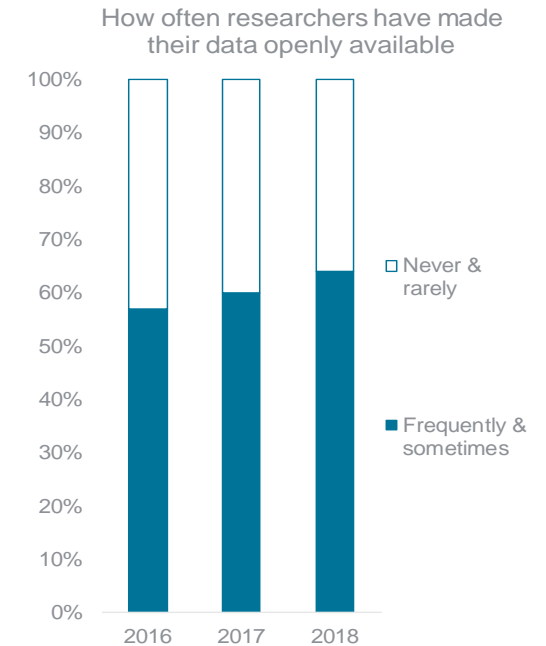
### 3) Researchers:

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

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%

# Increasing focus on requiring data

## Research data policy

---

EPJ strongly encourages that all datasets on which the conclusions of the paper rely should be available to readers. We encourage authors to ensure that their datasets are either deposited in publicly available repositories (where available and appropriate) or presented in the main manuscript or additional supporting files whenever possible. For further information on recommended repositories please see [List of Repositories](#)

5. **Data Underlying Published Research Results Will Be Accessible and Open Immediately.** The foundation will require that data underlying the published research results be immediately accessible and open. This too is subject to the transition period and a 12-month embargo may be applied.



# At Elsevier we are also implementing TOP

[About Elsevier](#) [Products & Solutions](#) [Services](#) [Shop & Discover](#)

[Home](#) > [Authors](#) > [Author resources](#) > [Research Data](#) > [Research Data Guidel...](#)

## Research Data Guidelines

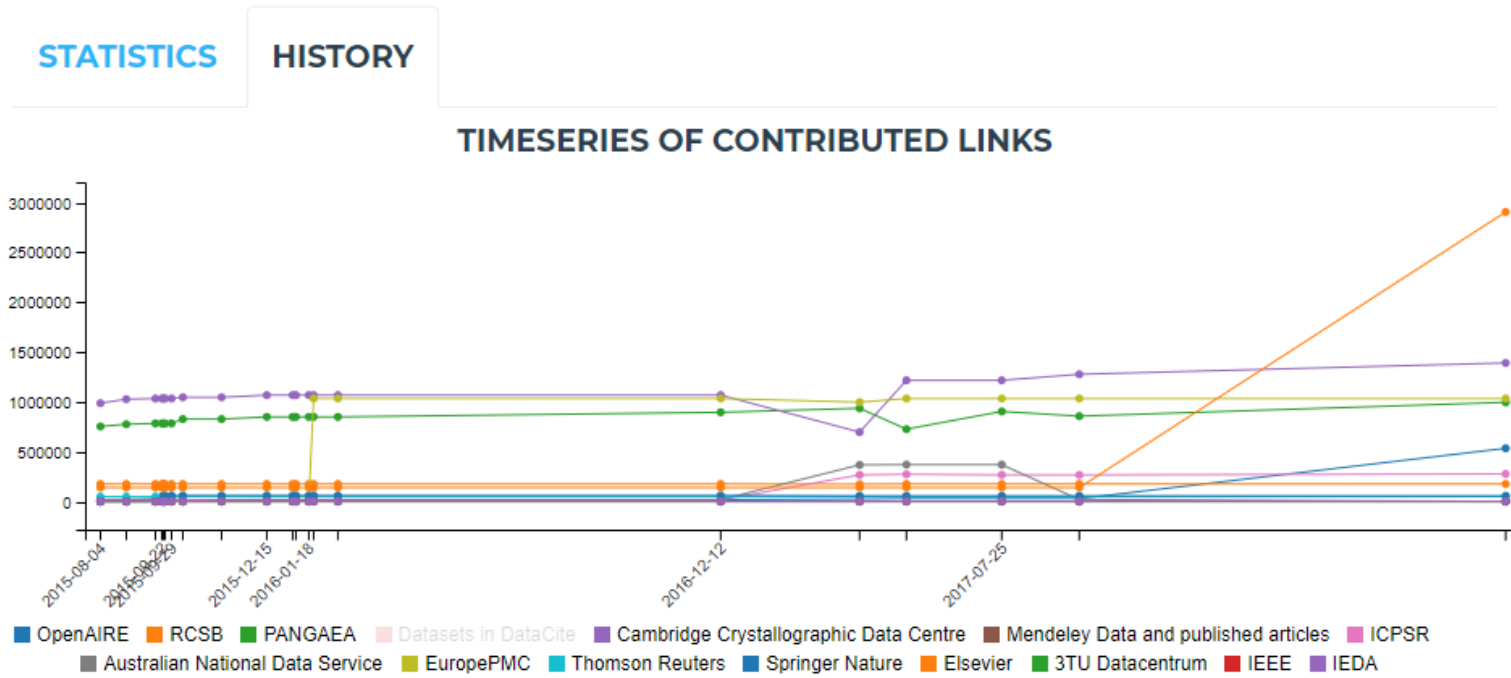
Data provides the core of your publication and is the foundation on which scientific, technical and medical knowledge is built. As a researcher, you are encouraged – and sometimes mandated – to store your research data and make it accessible and discoverable, so others can reuse it.

Elsevier journals have different guidelines when it comes to research data, either encouraging or requiring you take certain actions. Although much research data is disseminated as part of journal articles, a host of other data is not made available through article publication. These guidelines concern research data that often underlies, but exists outside of research articles. You can use the information below to understand the requirements of your chosen journal, so you can make informed choices about where and how to publish.

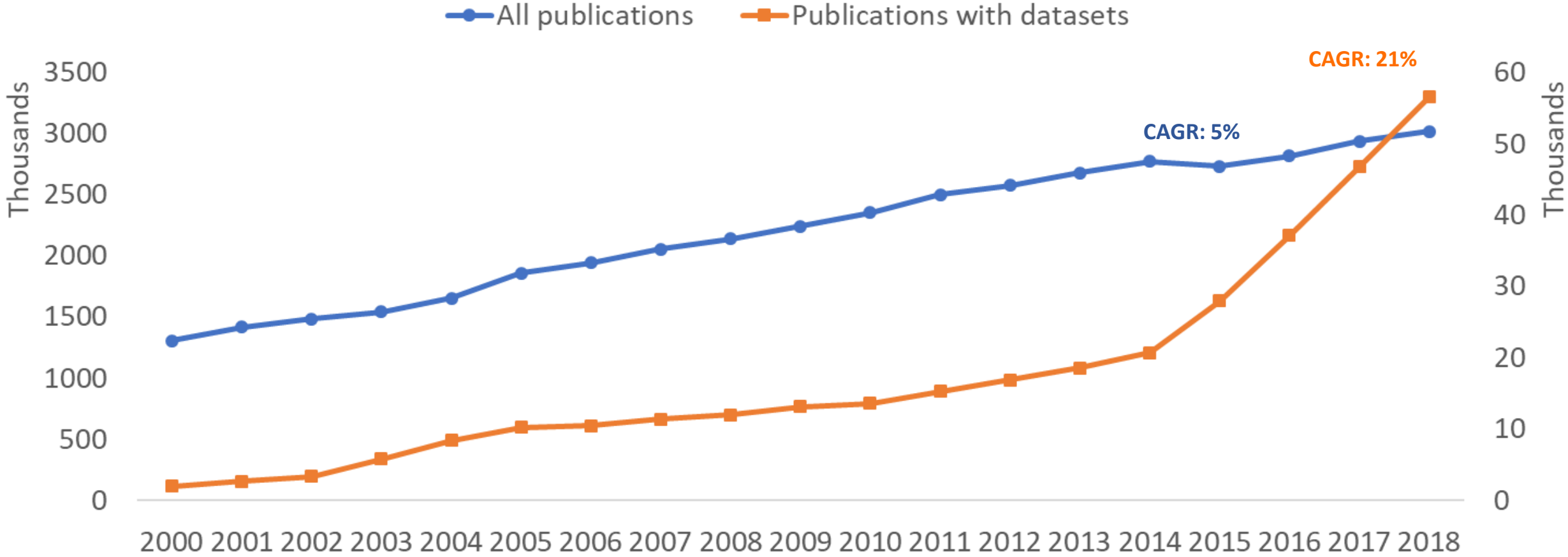


<https://www.elsevier.com/authors/author-resources/research-data/data-guidelines>

# Publishers now also starting to contribute in bulk

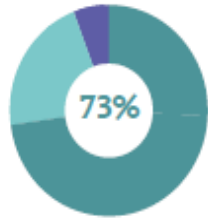


# Research Data Management adoption is growing very fast worldwide



Source: Scopus data as of July 9 2019, Scholix data extracted on May 5, 2019 – CAGR = Compound Annual Growth Rate

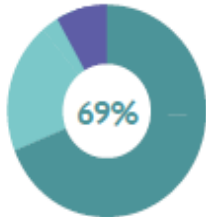
# Researcher motivation for sharing data



Having access to others' research data benefits/  
would benefit my own data



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



Sharing research data is important for doing  
research in my field

Which circumstances would motivate you to share your data?	2019 RANK
Increased impact and visibility of my research	1
Public benefit	2
Getting proper credit for sharing data	3
Journal/publisher requirement	4
Transparency and re-use	5
Funder requirement	6
Institution/organization requirement	7
Trust the person requesting my data	8
It was made easy and simple to do so	9
Freedom of information request	10
It was a field/industry expectation	11
Other (please specify)	12
I would never share my data	13
<b>TOTAL</b>	



# Sharing data works:

## *25% higher citation impact*

### The citation advantage of linking publications to research data

Giovanni Colavizza<sup>1,2,\*</sup>, Iain Hrynaszkiewicz<sup>3,4</sup>, Isla Staden<sup>1,5</sup>, Kirstie Whitaker<sup>1,6</sup>, Barbara McGillivray<sup>1,6</sup>

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](mailto: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.



# Articles with datasets outperform

y-axis

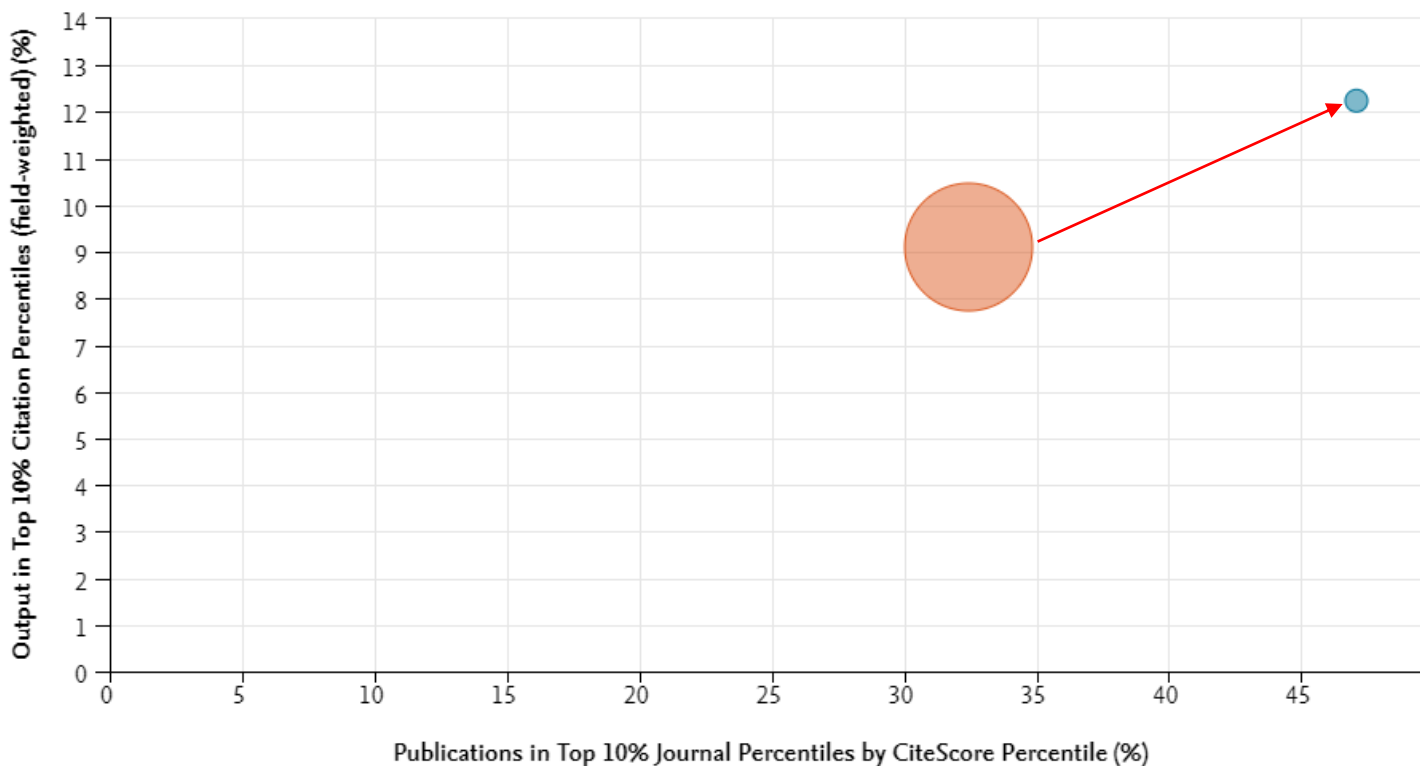
Output in Top 10% Citation Percentiles (field-weighted)

x-axis

Publications in Top 10% Journal Percentiles by CiteScore Percentile

Bubble size

Scholarly Output



Taiwan - Publications with Datasets

Taiwan

Hide all chart labels

View list of Scopus Sources for the selected Researchers and Groups

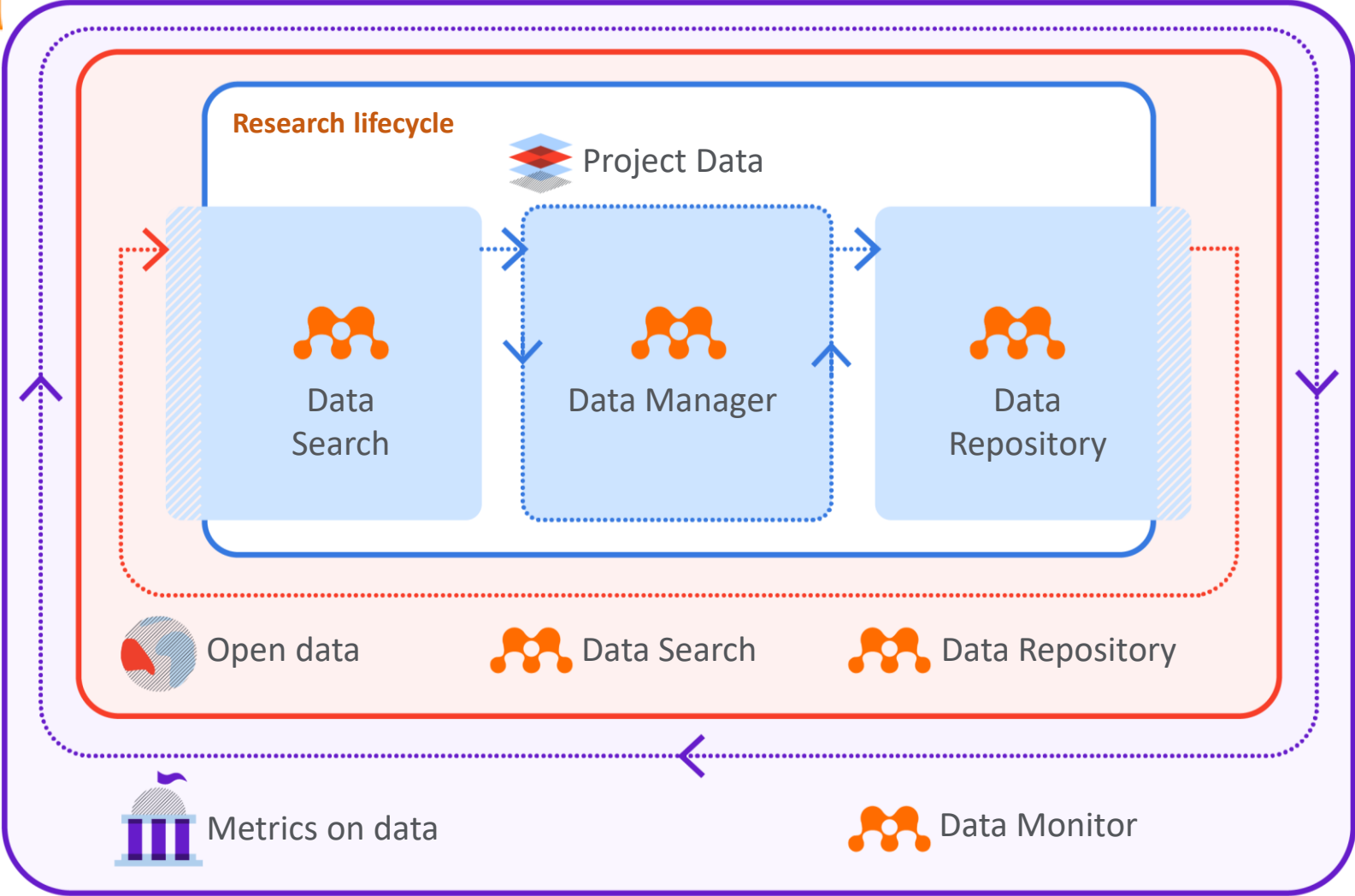


Scholarly Output



# Mendeley Data

A modular, cloud-based platform designed for institutions, to manage the entire lifecycle of research data



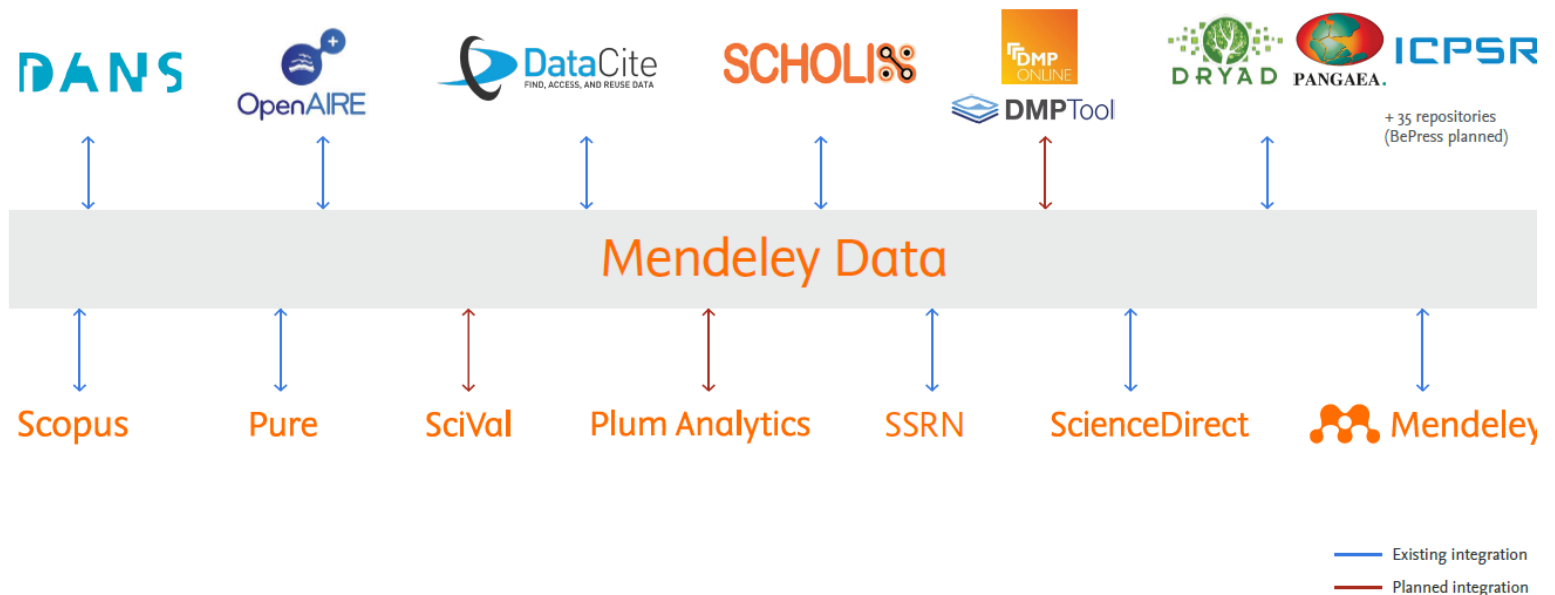
Mendeley Data



# How we deliver

- 1. Open system:** through open APIs, modules can be integrations with other RDM tools
- 2. Data remains private** at or owned by institution
- 3. System is integrated** with the researcher workflows, to ensure simple and clear use
- 4. Researchers continue to work the same way,** avoiding additional bureaucracy and administration

Mendeley Data already integrates through open APIs with the global Research Data Management ecosystem, as well as other Elsevier solutions

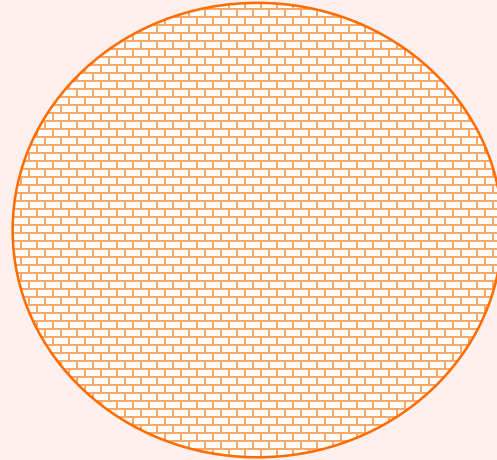


# The problem: finding the best way to keep track of datasets



## Internal datasets

For example: research data on an institutional repository



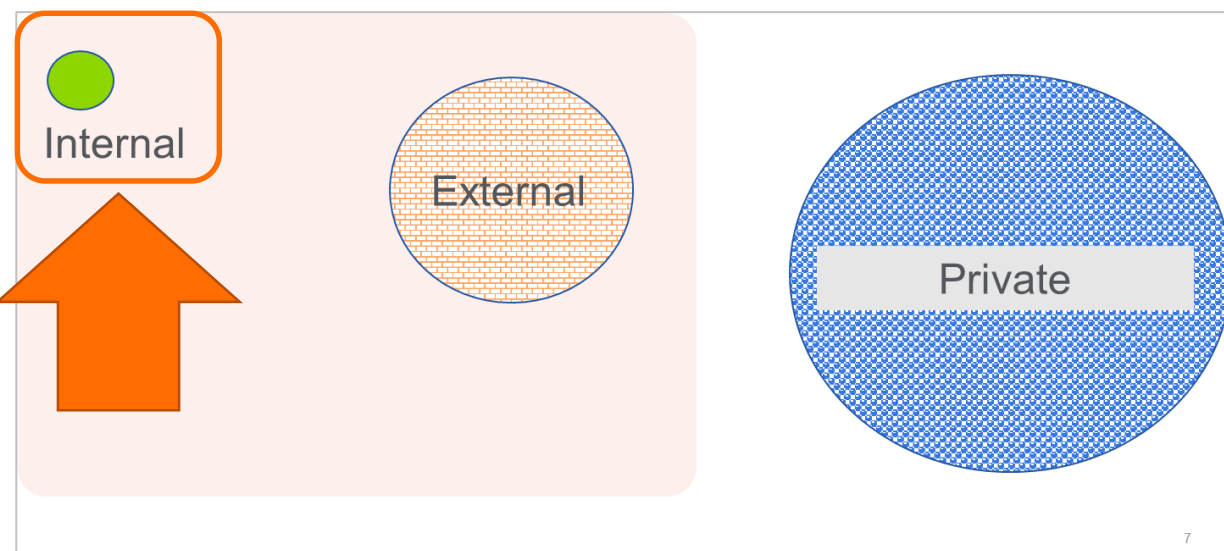
## External datasets

For example: research data on subject or domain repository



# Example: University of Virginia

LibraData data repository launched March 29, 2016: 149 datasets indexed as of 29 Sep 2019



The screenshot shows the UVA DataVerse website interface. At the top, the text "UVA DATAVERSE" is displayed in large orange letters. Below this, there are two main sections:

- LIBRADATA:** "LibraData: Uva's Scholarly Research Dataverse" with a logo.
- UNIVERSITY of VIRGINIA LIBRARY:** "Uva Library Dataverse" with a logo.

Below these sections is a search bar with the text "Search this dataverse..." and a "Find" button. To the right of the search bar is a link for "Advanced Search".

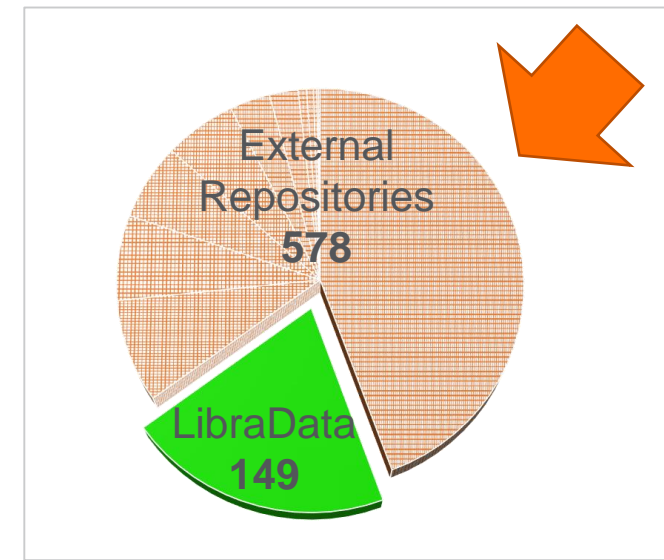
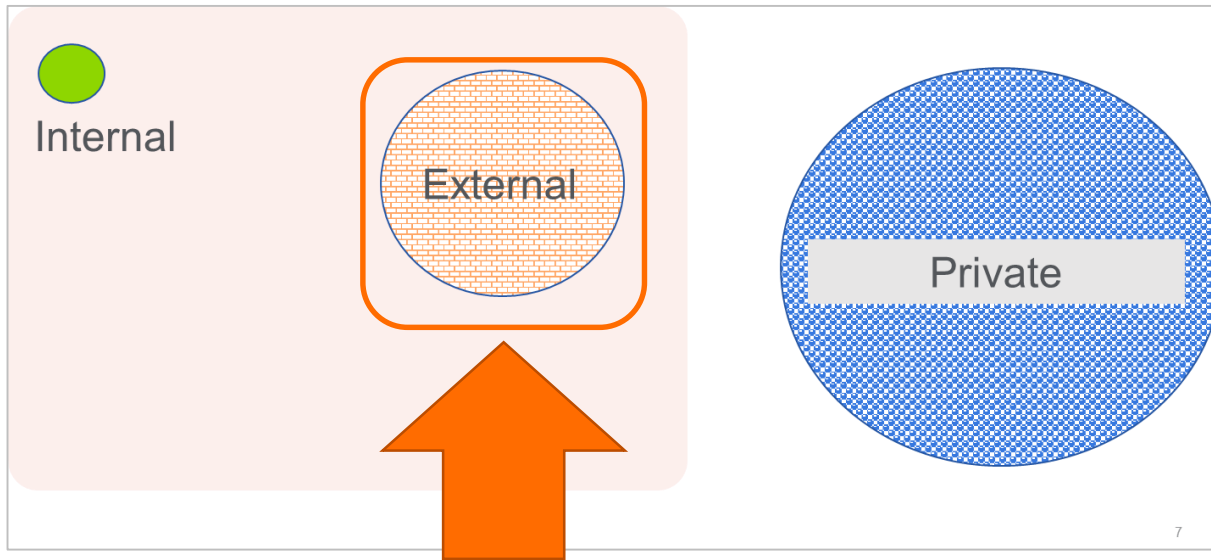
The search results section shows a list of items:

- Dataverses (22)**
- Datasets (149)** (This item is highlighted with an orange box and a large orange arrow points to it from the right.)
- Files (1,348)**

On the right side of the results, it says "1 to 10 of 171 Results". Below this, a preview of a result is shown: "3D Still Life - apple with bite - photogramr" dated "Sep 25, 2019 - Cultural Heritage".

Source: <https://dataverse.org/blog/uva-library-launches-%E2%80%99Libra-data%E2%80%9D%E2%80%94university-virginia-dataverse-repository>  
<https://dataverse.lib.virginia.edu/dataverse/uva>

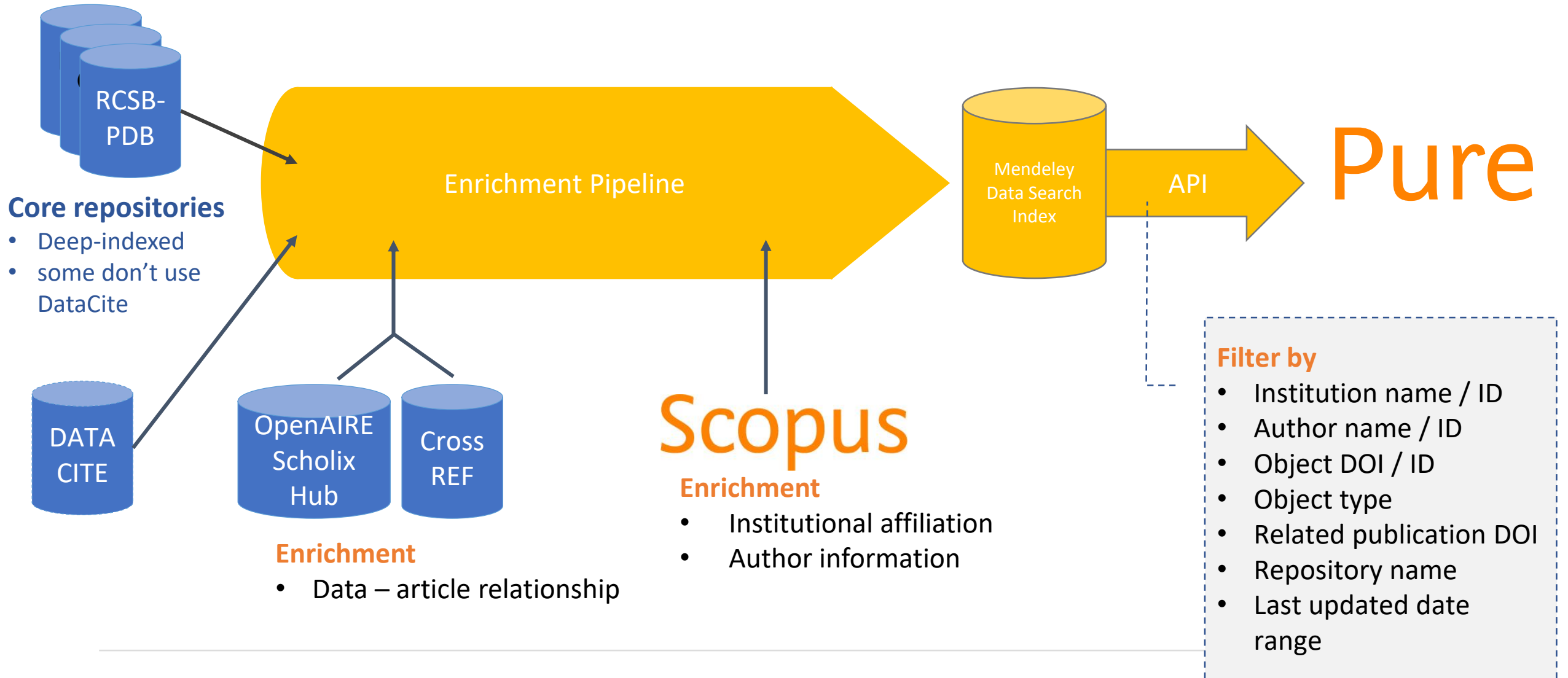
# At the University of Virginia, 80% of the datasets are published in 12 other repositories



This is not at all uncommon: we have found that in most universities more than 80% of the datasets are published in external data repositories



# Our solution: enrichment pipeline + API



## How do we deal with key requirement?

“I want to ingest data only into Pure” → we only ingest data repositories and also allow you to select the object type to ingest

# A flexible integration

## DataSearch integration

In this configuration page you can set up integration with DataSearch to automat

What is the DataSearch integration

Enabling the integration with DataSearch will automatically create Dataset conten

Enabling the DataSearch integration below will turn on the 'DataSearch: Import D

DataSearch integration

**Turn on** the DataSearch integration to source existing and future Dataset content.

**Turn off** the DataSearch integration to pause the integration and prevent new Dataset content being populated in Pure.

DataSearch integration

On

Institution IDs \*

Enter a comma-separated list of your institution's InstitutionalID and ScopusID to find Datasets from DataSearch.

Multiple institution IDs supported:  
Mendeley, SciVal, Scopus

Multiple data repositories supported

## DataSearch sources

DataSearch sources

MENDELEY\_DATA

A comma-separated list of sources to limit DataSe

Show available sources

The following sources are available through DataSearch

APOLLO  
ARRAYEXPRESS  
ARXIV  
AUSTRALIAN\_DATA\_ARCHIVE  
BIOLOGICAL\_MAGNETIC\_RESONANCE\_DATABANK  
CLINVAR  
COLLABORATIVE\_RESEARCH\_IN\_COMPUTATIONAL\_NEUROSCIENCE  
DATABRARY  
DATASPACE  
DRYAD  
DSPACEUNIVERSITYOFWASHINGTON  
FOURTUCCENTREOFRESEARCHDATA  
GBIF  
GEO  
GEOROC  
HARVARD\_DATAVERSE  
ICPSR  
LSHTMDATACOMPASS  
MEDECINSSANSFONTIERESFIELDRESEARCH  
MENDELEY\_DATA  
METPETDB  
NATIONAL\_INSTITUTE\_OF\_MENTAL\_HEALTH

+ Add content

Tasks

3 Datasets were found in DataSearch

2 names are each used on multiple Organisational units

## Import candidates from DataSearch for Some University

List of new Datasets imported from DataSearch for Some University

3 results

Dataset 1  
Jeffrey Lebowsky, 2019 DOI

Updated  
Found: 17 Sep 2019 12:43  
Update Source data | DOI

Dataset 2  
Dr. Strangelove, 2019 DOI

Updated  
Found: 17 Sep 2019 12:43  
Update Source data | DOI

Dataset 3  
Marty McFly, 2018 DOI

Found: 17 Sep 2019 12:43  
Import Source data | DOI

Non-relevant datasets can be rejected

Reject

# Allow your university repository to be discoverable by the world

**REPOSITORY TYPES**

- Article Repositories (5419475)
- Data Repositories (5352287)

**SOURCES**

- ScienceDirect (4863624)
- Zenodo (1273162)
- USGS Mineral Res. (995208)
- The Cambridge Crystallographic Data Centre (CCDC) (768013)
- arXiv (555851)
- GEOROC (478309)
- PANGAEA (388640)
- ClinVar (376028)
- Apollo Cambridge (231701)
- RCSB-PDB (138270)
- Gene Expression Omnibus (97540)
- NeuroMorpho (86892)
- PetDB (84938)
- DataSpace Princeton (75785)
- ArrayExpress (70654)
- NAVDAT (66781)
- Dryad (29228)
- Smithsonian (28457)
- DSpace Washington (24543)
- NeuroElectro (22726)

**Data for: Yield-power-law fluid propagation in water-saturated channel networks with application to rock grouting**  
Contributors: Liangchao Zou, Ulf Håkansson, Vladimir Cvetkovic  
Date: 2019-11-06  
... undefined

Top results from Data Repository sources. Show only results like these.

**Data for: English-medium instruction Calculus in Confucian-heritage culture classrooms: Why flipping is so hard?**  
Contributors: Natanael Karjanto, Lois Simon  
Date: 2019-11-06  
... Data for flipped classrooms in EMI-CHC.  
**Files:**  
Tabular Data

**Data for: Thermal Evolution in the Direct Chill casting of an Al-4 pct Cu Alloy using the Low-Reynolds Number Turbulence Model**  
Contributors: GAIUS NZEBUKA, Adekojo Waheed  
Date: 2019-11-06  
... Temperature, Dynamic viscosity, Velocity, solid fraction, kinetic energy etc  
**Files:**  
Tabular Data

**Data for: On the Modeling of a Contact-Aided Cross-Axis Flexural Pivot**  
Contributors: Pietro Bilancia, Larry Howell, Spencer Magleby, Giovanni Berselli  
Date: 2019-11-06

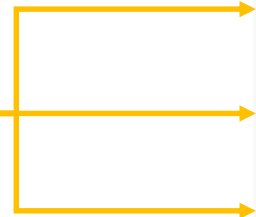
DSpace Washington (24543)

# Deep indexing capability

Find Research Data   My Datasets   New Dataset

influenza vaccination meta-analysis

Deep-indexing



**Neuro Electro** **Influenza vaccination** for immunocompromised patients: systematic review and **meta-analysis** from a public health policy perspective. ✕

Name	Matched	Outcome measure	Influenza subtype	Comparator	Number of studies	Pooled ES (95% CI)
Details	✓					
Table 4	✓	SC1	A(H1N1) (S)	VICT	50*	0.55 (0.43 to 0.71)
Table 5	✓					
Table 3	✓	SC1	A(H3N2)	VICT	47*	0.55 (0.41 to 0.73)
Table 1						
Table 2		SC1	B	VICT	44*	0.48 (0.36 to 0.62)

[Go to Data Source](#)

# Integration with Scopus

Scopus

1,338 document results

TITLE-ABS-KEY ( meta-analysis AND of AND influenza )

[Edit](#) [Save](#) [Set alert](#) [Set feed](#)

Search within results...

Refine results

[Limit to](#) [Exclude](#)

Access type ⓘ

- Open Access (425) >
- Other (113) >

[Documents](#) [Secondary documents](#) [Patents](#)

[View Mendeley Data \(24\)](#)

[Search your library](#)

Analyze search results

[Show all abstracts](#) Sort on: [Cited by \(highest\)](#)

All [CSV export](#) [Download](#) [View citation overview](#) [View cited by](#) [Save to list](#) [Print](#) [Email](#) [PDF](#)

	Document title	Authors	Year	Source	Cited by
<input type="checkbox"/> 1	Measuring inconsistency in meta-analyses	Higgins, J.P.T., Thompson, S.G., Deeks, J.J., Altman, D.G.	2003	British Medical Journal	24531

# Mendeley Data Search corpus of dataset

- Mendeley Data Search has a large number of datasets and it keeps on growing!
- To date, we have indexed **10.7M records** in **35+ multidisciplinary repositories**

Find Research Data My Datasets My Projects New Dataset FAQ

influenza vaccination

Filter Results [Reset](#) 507 results for influenza vaccination

FILES [^](#)

- Tabular Data (251)
- Text (144)
- Document (54)
- File Set (48)
- Software/Code (12)
- Unknown File Type (8)
- Slides (3)
- Image (3)

REPOSITORY TYPES [^](#)

- Data Repositories (507)

SOURCES [^](#)

- RCSB-PDB (178)
- Zenodo (86)
- Gene Expression Omnibus (77)
- ArrayExpress (63)
- NeuroElectro (74)

Stabilization of **influenza vaccine** enhances protection by microneedle delivery in the mouse skin. [×](#)

Name	Matched	Vaccines2	Week 1	Week 2	Week 4
Details	✓		IgG titers3 (×10 <sup>3</sup> )	IgG titers (×10 <sup>3</sup> )	IgG titers (×10 <sup>3</sup> )
Table 1	✓	PR8i	15.2±5	51.2±20	102.4±28
Table 2	✓	PR8i + Trehalose	12.8±3	60.8±25	90.4±20
		Mock	0.1±0.05	0.1±0.05	0.1±0.05

[Go to Data Source](#)



Prague  
2019

Pure  
International  
Conference

#prcn2019

Health-related behaviors and effectiveness of trivalent inactivated versus live attenuated **influenza vaccine** in preventing **influenza-like illness**

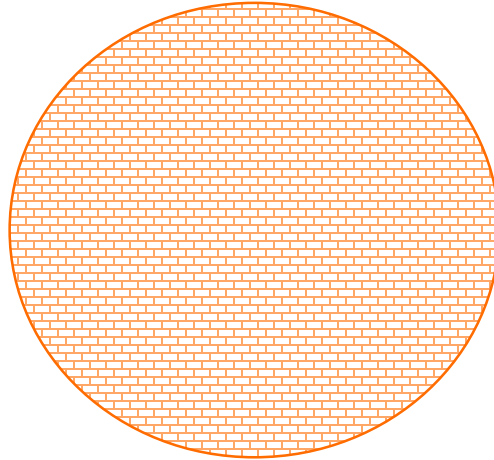


# The problem: helping researchers share data within their project when they collaborate



## Internal datasets

For example: research data on an institutional repository



## External datasets

For example: research data on subject or domain repository




## Private datasets

For example: data shared privately by researchers working on a project



# For example: Data Manager


 Alzheimer project 2  
Alzheimer project to investigate Western blot for dementia/alzheimer correlation with genes \_\_\_\_\_ FOR DEMO PURPOSES

---


### Data sources Group


[+ Add project data source](#)




**Spencer's Dropbox**

 Prototype clips

**Wouter's Dropbox**



 Alzheimer2

 **Blots**

-  western blot adult and fetal.png
-  western blot bacteria.tif
-  western blot brain









### Shared datasets Group

**Your shared datasets**


 Alzheimer's research: data analysis 

[+ Add a new shared dataset](#)

**Shared by others**

-  Adult and fetal western blot 
-  Alzheimer Research project 2 - shared data v2 
-  Western blots - round 2 
-  Western Blot - round 1 







### Published data Public



Nothing published

### Select a data source

Add a new data source

-  Dropbox
-  Box
-  Google Drive
-  Azure
-  DAV WebDAV
-  OneDrive

# RDM Librarian Academy Units



## Eight Online Units

1. Foundations of Research Data Management (RDM)
2. Navigating research data culture
3. Advocating and marketing the value of RDM in libraries
4. Establishing and managing data services in libraries
5. Project management (as relates to RDM services)
6. Overview of common tools (programming and tools, R)
7. Python, Jupyter Notebook
8. Overview of platform tools (OSF, Mendeley Data et. al.)

# Partner Institutions:

- Harvard Medical School
- Harvard University
- Simmons University
- Boston University
- Brown University
- Massachusetts College of Pharmacy and Health Science University
- Northeastern University
- Tufts University
- Elsevier



BROWN



ELSEVIER

# RDM Library Academy offers free online course for librarians and researchers



rdm library academy canvas



All

Maps

Images

Shopping

News

More

Settings

Tools

About 238,000 results (0.43 seconds)

## Research Data Management Librarian Academy - Canvas ...

<https://www.canvas.net> › [browse](#) › [simmonsu](#) › [courses](#) › [research-data-ma...](#) ▼

The Research Data Management **Librarian Academy** (RDMLA) is a free online ... advocating and marketing for **RDM** services in **libraries**, project management, ...

<https://www.canvas.net/browse/simmonsu/courses/research-data-management>



We would love to work  
with you

[adam.goh@elsevier.com](mailto:adam.goh@elsevier.com)

