



ELSEVIER

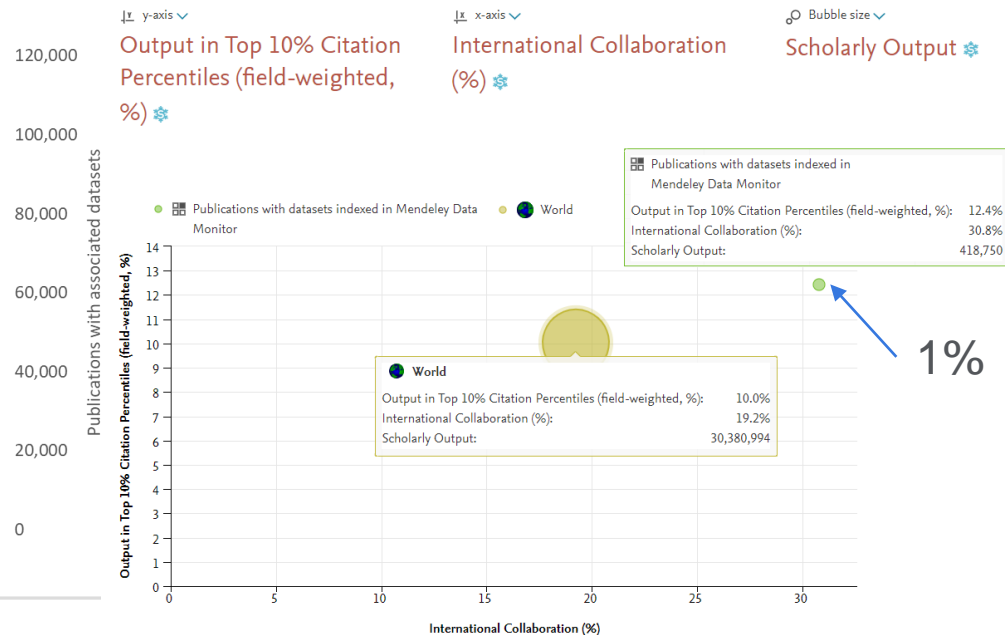
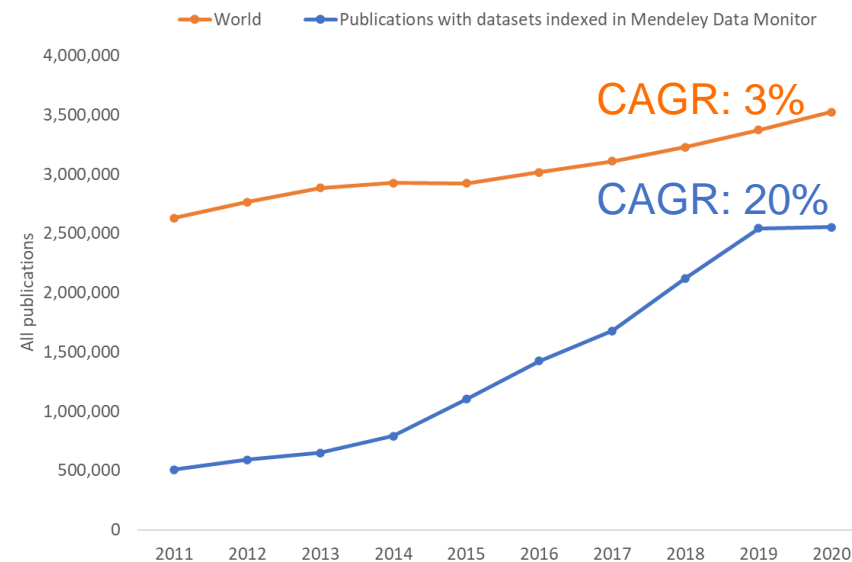
Session II: Exploring the current state of data citation methods and tools

Changing the Culture of Data
Management and Sharing
A Workshop
28 – 29 April 2021

Alberto Zigoni

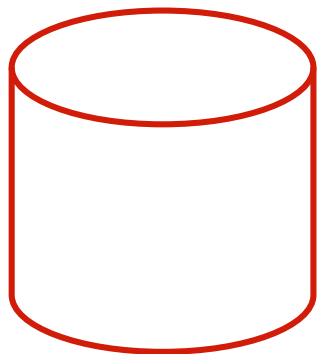
Market Development Director, Research Data Management Solutions, Elsevier BV

The practice of linking articles and datasets is growing and is associated with higher collaboration and citation impact



Source: SciVal, Data Monitor. Data extracted on April 21, 2021 - CAGR: Compound Annual Growth Rate

But where is your data?



~90%

of the public research data is hosted on external general subject or domain repositories

~10%

of public research data is hosted on institutional repositories



Dr. Andrew White
Director of Library
Information Services
at Rensselaer
Polytechnic Institute

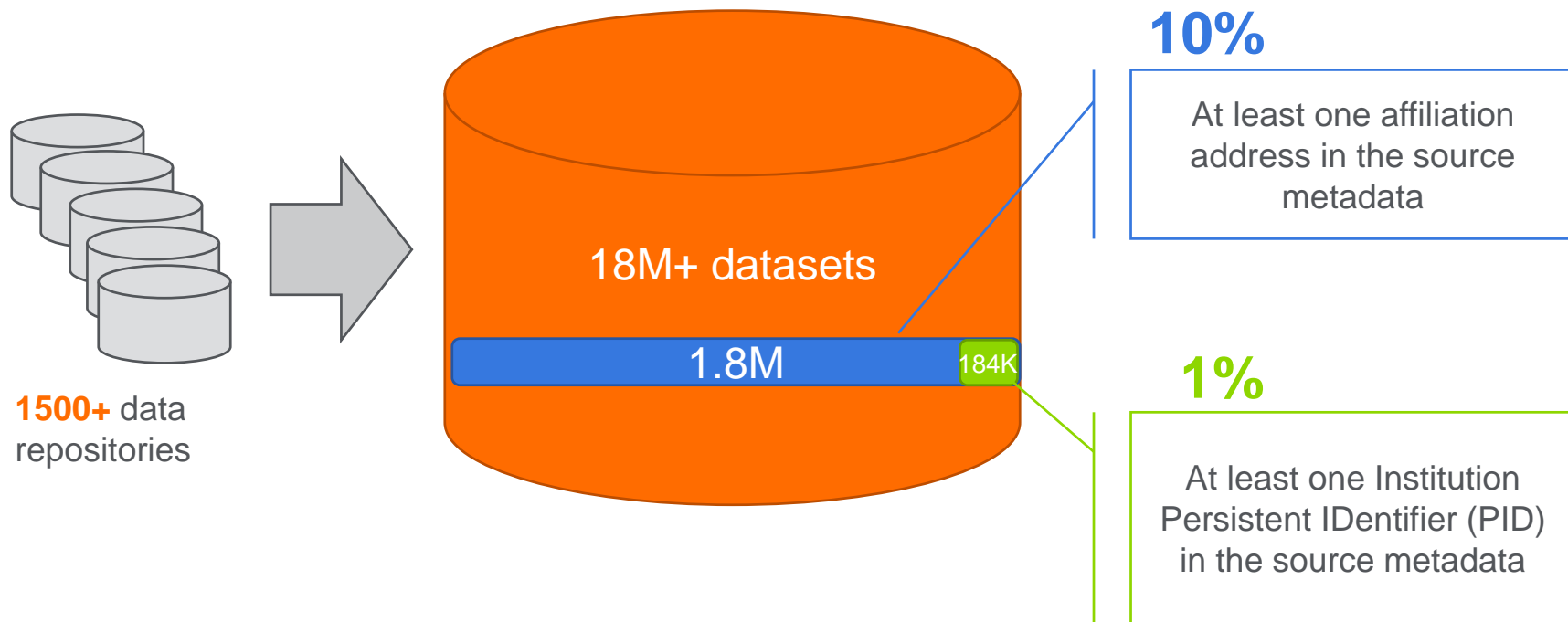
“

There is a broad spectrum of RDM practices used by our researchers, often due to a multitude of data types and a range of discipline-specific practices.


We see the libraries role as one that is focused on tracking archival practices of raw research data sets, assigning appropriate ontologies/metadata, providing appropriate linking and checking that the data has
”
Been stored in a secure location.





Why is it difficult to track datasets?



A great example of adoption of PIDs for institutions

 Mendeley Data

Find Research Data My Data Projects Collections Moderation Data Monitor Admin FAQ Alberto 

INSTITUTION_ID(01an7q238) → ROR ID of an institution 

[Advanced search help](#)

Filter Results **568 results**

DATA TYPES ^


- ☐ Dataset (561)
- ☐ Other (4)
- ☐ Document (2)
- ☐ Collection (1)

SOURCE TYPES ^


- ☒ Data Repositories (568)

SOURCES ^


- ☐ DRYAD (365)
- ☐ UC Berkeley (182)
- ☐ UC Davis (6)
- ☐ UC Merced (3)
- ☐ UC San Francisco (3)
- ☐ Utah State University (2)
- ☐ California Digital Library (1)
- ☐ UC Irvine (1)
- ☐ UC Los Angeles (1)
- ☐ UC Santa Barbara (1)
- ☐ UC San Diego (1)
- ☐ Stanford Digital Repository (1)
- ☐ University of Massachusetts (UMass) Amherst (1)

 **Data from: A test for rate-coupling of trophic and cranial evolutionary dynamics in New World bats**

Contributors: Shi, Jeff, Westeen, Erin, Rabosky, Daniel
Date: 2021-01-01
Source: DRYAD
01an7q238
Data Types:
Dataset
Export: [APA](#) [BibTeX](#) [DataCite](#) [RIS](#)

 **Why aren't rabbits and hares larger?**

Contributors: Tomiya, Susumu, Miller, Lauren K.
Date: 2021-01-01
Source: DRYAD
01an7q238
Data Types:
Dataset
Export: [APA](#) [BibTeX](#) [DataCite](#) [RIS](#)

 **Co-expression networks in Chlamydomonas reveal significant rhythmicity in batch cultures and empower gene function discovery**

Contributors: Salomé, Patrice, Merchant, Sabeeha

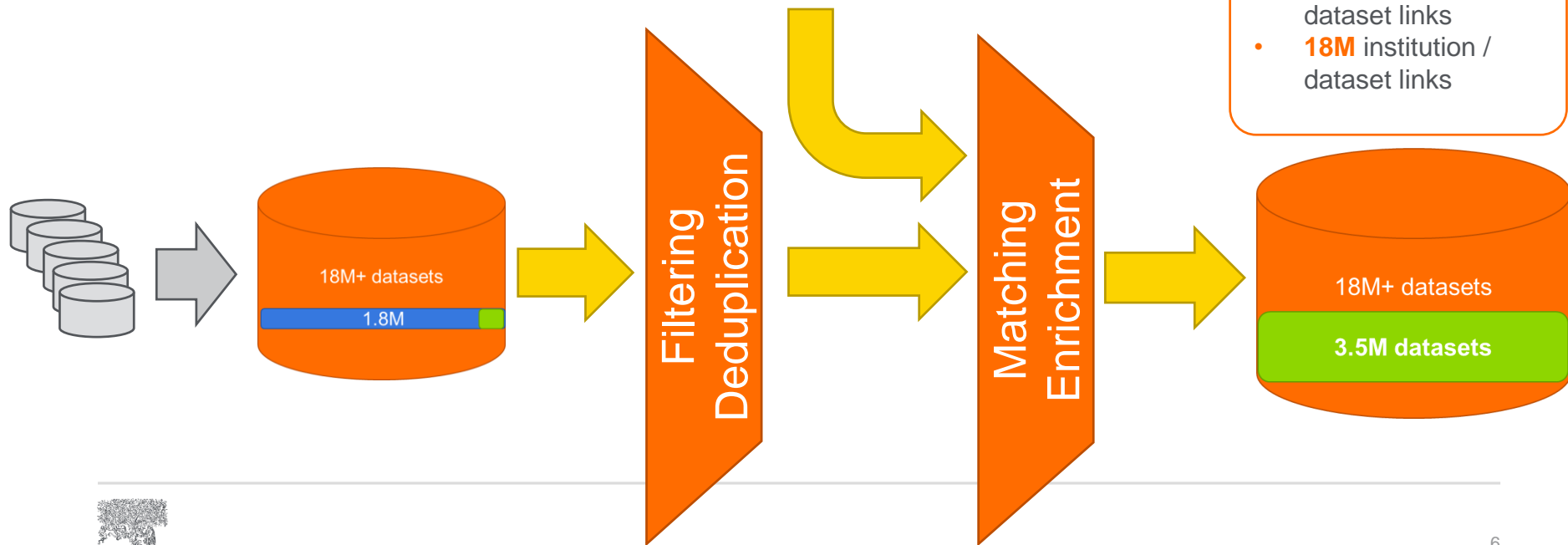
- Dryad has consistently started to use ROR identifiers for institutions
- Affiliation metadata is significantly better than other repositories
- However, there are lot of repositories where researchers publish their data...

Using Data Science to enrich dataset metadata

Scopus[®]
SciVal

- Publication metadata
- Author profiles
- Institution profiles

- **3.5M** datasets with at least one institution **PID**
- **14M** author / dataset links
- **18M** institution / dataset links



The results: 20x increase



Mendeley Data

[Find Research Data](#)

[My Data](#)

[Projects](#)

[Collections](#)

[Moderation](#)

[Data Monitor](#)

[Admin](#)

[FAQ](#)

[Alberto](#)



INSTITUTION_ID(01an7q238) OR INSTITUTION_ID(508259)

SciVal ID of the same institution



[Advanced search help](#)

Filter Results

10550 results

DATA TYPES

- ☐ Dataset (6424)
- ☐ Tabular Data (1353)
- ☐ Collection (1072)
- ☐ Document (1032)
- ☐ Image (516)
- ☐ Text (249)
- ☐ File Set (211)
- ☐ Other (135)
- ☐ Software/Code (74)
- ☐ Video (74)
- ☐ Sequencing Data (7)
- ☐ Geospatial Data (2)
- ☐ Slides (1)

SOURCE TYPES

- ☒ Data Repositories (10550)

SOURCES

- ☐ The Cambridge Structural Database (3660)
- ☐ figshare Academic Research System (2425)
- ☐ RCSB-PDB (1163)
- ☐ DRYAD (576)
- ☐ Zenodo (478)
- ☐ HEPData (476)
- ☐ Strasbourg Astronomical Data Center (452)
- ☐ ICPSR (359)
- ☐ figshare SAGE Publications (194)
- ☐ ArrayExpress (188)
- ☐ UC Berkeley (188)
- ☐ PANGAEA (125)
- ☐ Mendeley Data (89)
- ☐ Gene Expression Omnibus (46)
- ☐ GENETICS - Genetics Society of America (GSA) (15)
- ☐ G3 - Genetics Society of America (GSA) (10)
- ☐ GigaDB (9)
- ☐ Apollo Cambridge (8)
- ☐ UC Davis (7)
- ☐ Open Science Framework (6)
- ☐ Johns Hopkins University Library (6)
- ☐ UC San Diego (5)
- ☐ Harvard Dataverse (5)
- ☐ Movebank (5)
- ☐ UC Merced (4)
- ☐ UC San Francisco (4)
- ☐ Globus Services (3)
- ☐ Environmental Protection Agency (3)



Casa de Pueblo, Spain: Bridging the Divide Between Built and Moveable Heritage Conservators Through Social Media

Contributors: Sanchez, Alberto Sanchez

Date: 2021-01-01

Source: figshare Academic Research System

508259

Data Types:

Video

Export: [APA](#) [BibTeX](#) [DataCite](#) [RIS](#)



Phenotype data, genotype input files, and scripts related to adaptation in wild house mice from Western North America

Contributors: Ferris, Kathleen, Chavez, Andreas, Suzuki, Taichi, Phifer-Rixey, Megan, Bi, Ke et al

Date: 2021-01-01

Source: DRYAD

01an7q238

Data Types:

Dataset

Export: [APA](#) [BibTeX](#) [DataCite](#) [RIS](#)



CCDC 2054467: Experimental Crystal Structure Determination

Contributors: kai, liu

Conclusions

- Data sharing is growing, and it contributes to collaboration and academic impact
- Metadata make data count! Adding metadata, including affiliation information, contributes to dataset discoverability and attribution
- Persistent Identifiers, particularly for authors and institutions, are the best approach to attribute datasets and be rewarded for sharing them
- When PIDs are not available, or metadata is scarce, analytical approaches can help improve metadata quality and enable tracking, reporting and showcasing



Thank you

Alberto Zigoni

a.zigoni@elsevier.com

Market Development Director

Research Data Management Solutions

Elsevier BV

