Mendeley Data

Deep Dive

September 2020
Juan García Morgado, Solutions Sales Manager
j.garcia@elsevier.com
Two series webminar:

- Mendeley Data deep dive (September 2020)
The team:

Juan García, Solution Sales Manager, Research Data Management, j.garcia@elsevier.com

Annapaola Migani, Customer Consultant, Research Intelligence, a.migani@elsevier.com

Edward Wedel-Larsen, Account Manager Spain & Portugal, E.Wedel-Larsen@elsevier.com

Alvaro Ybarra, Account Manager Spain, a.ybarra@elsevier.com

Makumba Jawneh, Account Manager, m.jawneh@elsevier.com

Agne Karose, Customer Marketing Manager, a.baltrisiunaite@elsevier.com
Agenda

- The value of Research Data Management (RDM)
- Mendeley Data for Institutions
- Mendeley Data deep dive
The value of Research Data Management (RDM)
When talking about research data, we mean:

**All forms of data**
- Raw data
- Processed data
- Protocols, methods, workflows
- Machine & environment settings
- Scripts, analyses & algorithms

**FAIR Data Principles**
- Findable
- Accessible
- Interoperable
- Reusable
RDM adoption is growing very fast worldwide

Annual Growth: 21%

Annual growth: 5%

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

Sharing data works: **25%** higher citation impact

The citation advantage of linking publications to research data

Giovanni Colavizza¹,²,*, Iain Hrynaszkiewicz³,⁴, Isla Staden¹,⁵, Kirstie Whitaker¹,⁶, Barbara McGillivray¹,⁶

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

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.
The impact of sharing data in (for example) the US

Data sharing leads to:
- Higher citations
- Increased collaborations

Analysis in Trends module, selecting the specific country and institution; exported these articles as a separate set and benchmarked against overall institutional output in Benchmarking module.
Research area is based on a cross section between Scopus (affiliation), Scholix (article-data links), as stored in Mendeley Data Monitor database.
Mendeley Data
For institutions
An end-to-end RDM solution for institutions

Mendeley Data

Consisting of four modules, this open cloud-based platform helps institutions manage the entire lifecycle of research data and enables researchers to safely access and share information wherever they are.

Researchers can discover, collect and share research data.

Librarians and administrators can moderate, manage, reports and showcase research data output regardless of which data repository researchers use.
The Mendeley Data advantage

**Mendeley Data Search**
A comprehensive data search engine including 20+ million datasets indexed from 1000s of data repositories

**Mendeley Data Repository**
A repository specialized for research data, to store and share datasets following the FAIR data principles

**Mendeley Data Manager**
A collaborative team workspace where researchers can share project data, integrate external data sources securely and prepare datasets for publication

**Mendeley Data Monitor**
A tool to track datasets published by researchers both within and outside the institution, to facilitate compliance with funders' mandates and enable reporting and showcasing

The Mendeley Data platform gives you full control to successfully drive forward your institutional Research Data Management (RDM) policy
Mendeley Data free vs. paid at a glance

- **Free**
  - Mendeley Data Search
  - Create dataset up to 10GB
  - Mint DOI

- **Paid**
  - Mendeley Data Repository
  - Mendeley Data Manager
  - Mendeley Data Monitor

---

**Institution**

**Researcher**
# Assessing RDM for an institution

<table>
<thead>
<tr>
<th>Research data location</th>
<th>Purpose</th>
<th>Considerations</th>
</tr>
</thead>
</table>
| Stored on your *institutional data repository* | • A place for researchers to store and share their data easily in a trusted and secure place  
• Accounts for ~10% of public research data | • Do you have an institutional data repository or are you planning to get one?  
• If you use a generalist repository, does it work well enough for the differing needs for sharing data?  
• How do you showcase data in your institutional repository? |
| Privately shared on university infrastructure | • Accounts for ‘active’ data that researchers work on while doing research  
• Strong data management means that institutions help their researchers in this active phase, and help make data more FAIR | • Do you have a data policy?  
• Do you require researchers to have data management plans, and if so for compliance only?  
• Do you have Data Stewards or Champions?  
• How do you help researchers manage and collaborate with others on active data?  
• Do you have visibility of active data?  
• If a researcher leaves how do you ensure that their data is understandable and reusable by others? |
| Publicly shared on *generalist or domain repositories* | • Accounts for ~90% of public research data  
• It is considered good practice to encourage researchers to share data here | • Do you track, report, and get credit for this data?  
• Can your researchers find this data? How? |
# Mendeley Data

<table>
<thead>
<tr>
<th>Research data</th>
<th>Open Science (free layer)</th>
<th>Mendeley Data for institutions</th>
</tr>
</thead>
<tbody>
<tr>
<td>On your institutional data repository</td>
<td></td>
<td>Mendeley Data Repository</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Large size data &amp; local storage</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Use your institutional DOI</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Institutional branding and showcasing</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Data collections</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Administration, moderation, curation, custom metadata</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Article-data linking</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Automated integration with reporting systems (e.g. CRIS)</td>
</tr>
<tr>
<td>Active data privately shared on institution infrastructure</td>
<td></td>
<td>Mendeley Data Manager</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Link with active data</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Integrates with generalist cloud storage provider tools</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Collaborate on projects</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Ability for Data Librarians or Data Stewards to support in active phase</td>
</tr>
<tr>
<td>Publicly shared on generalist or domain repositories</td>
<td>Mendeley Data Repository</td>
<td>Mendeley Data Monitor &amp; Mendeley Data Search</td>
</tr>
<tr>
<td></td>
<td>• Create datasets up to 10GB</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• DOI</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Data versioning</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mendeley Data Search</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Find research data in over 2000 data repositories</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Track and report on your data found in over 2000 data repositories</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• See how other institutions manage their data</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Your data repository is deep indexed in data search</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Automated integration with your reporting systems (e.g. CRIS)</td>
</tr>
</tbody>
</table>
Mendeley Data Repository

- Mendeley Data indexed by OpenAIRE index
- OpenAire Zenodo repository indexed by Mendeley Data Search
- Mint DOIs for Mendeley Data Repository
- Data Cite indexed by Mendeley Data Search
- Contributed by Mendeley Data to Scholix
- Indexed by Mendeley Data Search and Data Monitor
- Consumed by Scopus and ScienceDirect
- Notify new articles to Monitor for data sharing compliance
- Datasets appear as records on Scopus (planned)
- Mendeley Data Repository datasets are automatically synced with the Pure curation workflow
- Projects, grants, equipment, showcase on portal (planned)
- Mendeley Data usage is accessible through Plum API and widget
- Plumx metrics (citations, usage, social mentions) are captured and shown on Mendeley Data Repository
- Publish datasets alongside an article within the SSRN publication flow
- Publish or link datasets alongside an article on ScienceDirect within the ScienceDirect publication flow
- User identity & login
- Library (planned)
- Notes (planned)
- Projects (planned)

- For more than 35 repositories the metadata as well as the underlying datasets are indexed by Mendeley Data Search
- First repositories are actively integrating with the free and open ‘push API’ of Mendeley Data Search
- Data Cite indexed by Mendeley Data Search
- Mint DOIs for Mendeley Data Repository
- OpenAire Zenodo repository indexed by Mendeley Data Search
- Mendeley Data indexed by OpenAIRE index
- Existing integration
- Planned integration
Mendeley Data Deep Dive
Mendeley Data Search

A powerful search engine specialized for data
Mendeley Data Search indexes 21+ million datasets from 1,700+ domain specific, generalist and institutional repositories

Data search accuracy through deep indexing and advanced search
Mendeley Data Search has the power to index both the metadata and actual data within a dataset. This, combined with advanced query processing and search capabilities, ensures that researchers get more accurate and comprehensive search results.
An intuitive and powerful search experience

Mendeley Data Search enables researchers to:

• Use advanced search query syntax to build sophisticated queries
• Leverage state-of-the-art query processing capabilities (lemmatization, concept search, acronym expansion) to improve the comprehensiveness of results
• Get more accurate results by leveraging deep-indexing capabilities
• Review the appropriateness of datasets in a glance, thanks to preview capabilities

Benefit for researchers: the data discovery section of the research lifecycle is more effective as the search engine is specialized for data and efficiently integrated with Scopus
An open API to fully integrate search within current workflows

Mendeley Data Search provides a powerful API to:
• Integrate search capabilities into third party tools used by your institution
• Index your existing institutional data repository within Data Search

Benefit for researchers: researchers can incorporate search results into their data processing pipelines

Benefit for the institution: institutions can integrate search results in their discovery platforms through the search API, as well as having their data repository deep-indexed by Data Search via the Push API.
Mendeley Data Repository

**Specialized for research data**
Equipping your researchers with a secure and trusted repository to store and share datasets following the FAIR data principles

**Gives full control to the Institution**
Enabling you to successfully showcase research data outputs and drive the implementation of your RDM policy

Findable ☑️, Accessible 🧐, Interoperable 🧮, Reusable 🔄

**CORE TRUST SEALS**

16 open data licenses

24.09.20
Custom metadata allows your institution and project teams to define metadata schema, for easy discoverability and reuse. You can specify additional fields to be captured for every dataset and you can use these fields any way you wish.

**Benefit for the library:** consistent categorization using custom metadata fields enables **better cataloguing of data**

**Benefit for researchers:** datasets have richer descriptive details, therefore increasing **discoverability and enhanced reusability** thanks to more accurate interpretation.

<table>
<thead>
<tr>
<th>Publishing</th>
<th>Metadata</th>
<th>Activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fields required or added by your library:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grant #:</td>
<td>Enter value</td>
<td></td>
</tr>
<tr>
<td>Scientific Discipline:</td>
<td>Enter value</td>
<td></td>
</tr>
<tr>
<td>Sponsorship:</td>
<td>Enter value</td>
<td></td>
</tr>
<tr>
<td>Faculty admin:</td>
<td>Georgina Rudd</td>
<td></td>
</tr>
</tbody>
</table>

Ex. of metadata input fields on a dataset, where the fields have been set by the author’s institution.
Moderate datasets to **maintain quality outputs**

As a moderator you can:

- **Review** every dataset prior to go-live
- **Make correction and additions**: edit dataset title and descriptions
- **Approve or return** the dataset to the author with comments to improve

Reviewing a dataset before publication
Organize datasets to **better showcase outputs**

Create and publish collections of datasets on your institutional showcasing page

- Share data on a particular theme or topic
- Include datasets from any repository your researchers use
- Make your datasets fully visible and searchable from your showcase page
Mendeley Data Manager

A collaborative research data management space
An online research project space for your researchers to use to manage data within a research project, wherever the team members are located.

Integrates with existing workflows
Researchers can use their current tools to collect project data and then share this seamlessly with project members.
A project space to foster collaboration and active data management

Data Manager allows researchers to:
• Collaborate on data regardless of their physical location and institutional affiliation
• Prepare to share research data in accordance with the FAIR data principles

Benefit for the institution: track research data activities and encourage the adoption of RDM best practices deep with the research lifecycle.

Benefit for researchers: facilitate collaboration with researchers globally and minimize the burden of RDM-related activities.

Project collaborators can easily be searched by email address and added to a project.
Access to project data is **seamless and controlled**

Researchers use a variety of tools to store project data, depending on personal preferences and institutional choices.

Data Manager enables researchers to integrate multiple data sources within a project, with controlled access and seamless transition of data into shared datasets.

**Benefit for the institution:** minimize the risk of data loss and support research integrity

**Benefit for researchers:** streamline RDM activities and easily onboard new project members, wherever they’re based.

External data sources can be referenced into projects. Never lose track of research data anymore!
Mendeley Data Monitor

**Puts institutions in the RDM driving seat:**

- Safeguard compliance with funder mandates and institutional RDM policy
- Maintain an overview of where research data is published
- Showcase Institutional research data output
Organize, curate and report on public research data to enhance visibility and compliance

It’s difficult for institutions to have full visibility of thousands of open data repositories and monitor which ones their researchers are using.

~90% of public research data is hosted on external generalist or domain specific repositories

~10% of public research data is hosted on institutional repositories

Our solution:

• Provides you with visibility for the entire research data output of your institution

• Integrates more than 20 million research data records from 1700+ generalist and domain specific repositories
Showcase public datasets to improve discoverability and reuse

As an institution you can:

- Get higher scores for grant applications by promoting public research data output of your Institution
- Improve your institutional research data visibility by making it easy to find and reuse
Compare and benchmark your research data output

As an institution you can compare and benchmark your research data output with other institutions

• via advanced data search queries,
  or
• by using Elsevier products featuring Mendeley Data Monitor integrations (e.g. SciVal)
Integrate with the Data Monitor API and apply **standard approval processes**

As an institution you can:

- Continue using your standard workflows and approval processes

**Integration with existing workflows**

Librarians can make use of their current tools to organize, curate and report on research data (e.g.: via CRIS or any other institutional system)
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 your 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
Wrap up

- The value of Research Data Management (RDM)
- Mendeley Data for Institutions
- Mendeley Data deep dive
Questions?
Thank you

Juan García Morgado
Solutions Sales Manager
j.garcia@elsevier.com