Mendeley Data
From disconnected datasets to relevant and complementary data:
The FAIR Data Principles
Two series webinar:

• From disconnected datasets to relevant and complementary data: The FAIR Data Principles (15th October)
• Mendeley Data deep dive (5th November)
The team:

- Christian Palazzesi (Account Manager) - IT
- Marina D’Ambrossio (Account Manager) – IT
- Rosella Davi (Account Manager) - NL
- Luigi Ruco (Solutions Sales Manager) – IT
- Agne Karose (Marketing Manager) – NL
- Juan García Morgado (Solutions Sales Manager) - SP
Agenda

• The value of Research Data Management (RDM)
• Mendeley Data for Institutions
• Mendeley Data and the FAIR principles
The value of Research Data Management (RDM)
When talking about research data, we mean:

- 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: 5%

Annual Growth: 21%

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

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

* 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 institution 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

Paid

Mendeley Data Repository

Mendeley Data Manager

Mendeley Data Monitor

Institution

Researcher

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

1. Private data
   - Data Search
   - Data Manager
   - Data Repository

2. Public data
   - Data Repository
   - Data Search

3. Metrics on data
   - Data Monitor
Integrated across the global RDM ecosystem

Long-term preservation of published datasets
- Mendeley Data indexed by OpenAIRE index
- OpenAire Zenodo repository indexed by Mendeley Data Search

Links between articles and datasets:
- Contributed by Mendeley Data to Scholix
- Indexed by Mendeley Data Search and Data Monitor
- Consumed by Scopus and ScienceDirect

Integrate with machine readable data management plans
- 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

Publish datasets alongside an article on Mendeley Data Repository
- Mint DOIs for Mendeley Data Repository
- Data Cite indexed by Mendeley Data Search

Publish or link datasets alongside an article on ScienceDirect publication flow
- Scopus
- Mint DOIs for Mendeley Data Repository
- Data Cite indexed by Mendeley Data Search

Plum Analytics
- Mendeley Data usage is accessible through Plum API and widget
- Plumx metrics (citations, usage, social mentions) are captured and shown on Mendeley Data Repository

SSRN
- Publish datasets alongside an article on Mendeley Data within the SSRN publication flow

ScienceDirect
- Publish or link datasets alongside an article on Mendeley Data within the ScienceDirect publication flow

Mendeley
- User identity & login
- Library (planned)
- Notes (planned)
- Projects (planned)

Scopus
- Mendeley Data Search results are visible on Scopus
- Notify new articles to Monitor for data sharing compliance
- Datasets appear as records on Scopus (planned)

Pure
- Mendeley Data Repository datasets are automatically synced with the Pure curation workflow
- Projects, grants, equipment, showcase on portal (planned)

SciVal
- Researcher and Institutional Dataset metrics

Plum Analytics
- Mendeley Data usage is accessible through Plum API and widget
- Plumx metrics (citations, usage, social mentions) are captured and shown on Mendeley Data Repository

SSRN
- Publish datasets alongside an article on Mendeley Data within the SSRN publication flow

ScienceDirect
- Publish or link datasets alongside an article on Mendeley Data within the ScienceDirect publication flow

Mendeley
- User identity & login
- Library (planned)
- Notes (planned)
- Projects (planned)

Scopus
- Mendeley Data Search results are visible on Scopus
- Notify new articles to Monitor for data sharing compliance
- Datasets appear as records on Scopus (planned)

Pure
- Mendeley Data Repository datasets are automatically synced with the Pure curation workflow
- Projects, grants, equipment, showcase on portal (planned)

SciVal
- Researcher and Institutional Dataset metrics

Plum Analytics
- Mendeley Data usage is accessible through Plum API and widget
- Plumx metrics (citations, usage, social mentions) are captured and shown on Mendeley Data Repository

SSRN
- Publish datasets alongside an article on Mendeley Data within the SSRN publication flow

ScienceDirect
- Publish or link datasets alongside an article on Mendeley Data within the ScienceDirect publication flow

Mendeley
- User identity & login
- Library (planned)
- Notes (planned)
- Projects (planned)

Scopus
- Mendeley Data Search results are visible on Scopus
- Notify new articles to Monitor for data sharing compliance
- Datasets appear as records on Scopus (planned)

Pure
- Mendeley Data Repository datasets are automatically synced with the Pure curation workflow
- Projects, grants, equipment, showcase on portal (planned)

SciVal
- Researcher and Institutional Dataset metrics

Plum Analytics
- Mendeley Data usage is accessible through Plum API and widget
- Plumx metrics (citations, usage, social mentions) are captured and shown on Mendeley Data Repository

SSRN
- Publish datasets alongside an article on Mendeley Data within the SSRN publication flow

ScienceDirect
- Publish or link datasets alongside an article on Mendeley Data within the ScienceDirect publication flow

Mendeley
- User identity & login
- Library (planned)
- Notes (planned)
- Projects (planned)
Mendeley Data and the FAIR principles
FAIR data principles

Data and supplementary materials have sufficiently rich metadata and a unique and persistent identifier: FINDABLE

Metadata and data are understandable to humans and machines. Data is deposited in a trusted repository: ACCESSIBLE

Metadata use a formal, accessible, shared, and broadly applicable language for knowledge representation: INTEROPERABLE

Data and collections have a clear usage licenses and provide accurate information on provenance: REUSABLE

Making data FAIR with Mendeley Data

Creating research data standards

Elsevier is a founding member of Force11, a community dedicated to improving knowledge creation and sharing. Working with other key stakeholders, we helped develop the FAIR Data Principles to support improved Findability, Accessibility, Interoperability and Reuse.

Our commitment to making data effective

Elsevier is committed to making data effective and developing better research data management processes and systems to support data sharing. The Mendeley Data platform helps researchers discover, collect, and share research data with the FAIR Data Principles at the core of the solution.

“One of the six main recommendations from the Force11 Manifesto was to treat data, software and workflows as a first-class citizen. Groups such as Enabling FAIR Data helped set the standard for making data FAIR across a wide range of stakeholders, including all of the leading publishers in the Earth Sciences.

Anita de Waard, VP Research Data Collaborations at Elsevier and FORCE11 founding member
Mendeley Data and the FAIR Data principles

Making data findable

• Datasets are indexed with metadata in common search indexes, such as Google Dataset Search via schema.org, DataCite Search, OpenAIRE with OAI-PMH, and Share from Open Science Framework

• Mendeley Data Search is an open search engine that indexes over 20 million datasets from thousands of public repositories—and Mendeley Data datasets include deep-indexing of both metadata and files

• Industry-leading advanced search functionality provides access to over 20 million datasets from thousands of data repositories—with state-of-the-art retrieval techniques to improve precision and recall

• Digital object identifiers (DOI) are assigned to all datasets in Mendeley Data Repository, as well as the underlying assets and versions.
Mendeley Data and the FAIR Data principles

Making data accessible

- All data held within Mendeley Data Repository remains owned and controlled by the researcher or institution, with access to 16 open data licenses should the owner decide to share the data publicly.
- Mendeley Data Repository is a data storage solution that ensures dataset owners retain control over access levels, with options to share data openly, place under embargo, or share privately within a controlled project environment.
- Open APIs at the Mendeley Data platform level allow all four modules to be used together, work as standalone modules, or integrate with any other RDM tool.
- Mendeley Data Monitor provides tool for institutions to track datasets published by researchers both within and outside the institution, to facilitate compliance with funders’ mandates and enable reporting and showcasing of research output.
Mendeley Data and the FAIR Data principles

Making data **interoperable**

- Allows code to run to reuse datasets, so they can be used for AI training such as image classification
- Integrates with other RDM tools to push and pull datasets from the repository via REST APIs using the JSON format
- Enables datasets to be updated with new versions for future interoperability while preserving provenance
- Offers controlled vocabularies and identifiers for standard fields and custom metadata fields

### Publishing Metadata Activity

**Fields required or added by your library:**

- Grant #: Enter value
- Scientific Discipline: Enter value
- Sponsorship: Enter value
- Faculty admin: Georgina Rudd

Ex. of metadata input fields on a dataset, where the fields have been set by the author’s institution
Supports standard metadata schema such as Dublin Core and schema.org, as well as the use of controlled vocabularies for standard fields and custom metadata fields.

Custom metadata fields can be configured to use values from existing taxonomies, for easier discoverability and reuse.

Allows institutions and researchers to add domain-specific custom metadata fields to datasets.

Encourages researchers to include step-by-step reproducibility guidance within the dataset description.
Support the FAIR data principles in practice

Case Study
Standardizing author citation metrics
dx.doi.org/10.17632/btxchxktyzw.1

Researchers used Mendeley Data Repository to publish a dataset of author citation metrics best practices. 100,000 scientists in 22 scientific fields and 176 sub-fields were included in the dataset, which standardized citations, h-index, co-authorship adjusted hm-index, citations based on different authorship positions, and a composite indicator. The dataset authors successfully applied the FAIR Data Principles in multiple ways:

- Dataset viewed over 100,000 times with 37,000 download
- Uses clear metadata descriptors and labeling
- Includes detailed reproducibility guidelines
- Folders used to separate code from data
Support the FAIR data principles in practice

Case Study
Reusing datasets to train deep-learning algorithms

Deep learning, one of the most data-intensive research fields, requires reusing high-quality data to train algorithms effectively. Since large amounts of data are involved, it’s critical that data retrieval and processing are as automated as possible, and that data is easy to integrate into existing workflows.

Google’s TensorFlow is the leading software framework for deep learning applications. Currently, it recommends four datasets from Mendeley Data Repository, including the plant leaves dataset, in the list of recommended datasets for image recognition algorithms—one of the most successful deep learning applications. These datasets support the FAIR Data Principles in multiple ways:

- The datasets have over 10,000 downloads, with Mendeley Data Repository supporting standard metadata discoverability protocols and advanced SEO techniques to ensure dataset discoverability.
- CC-BY licensed datasets authorize researchers to use the data.
- Interoperable formats integrate into existing content processing pipelines.
- High-quality, annotated data ensures algorithms are trained effectively.
- Datasets are accessible directly from Python code, including instructions for citing datasets within the code.

Visit our website to find out more about how Mendeley Data supports FAIR:
elsevier.com/solutions/mendeley-data-platform/fair
# Five facts about RDM from Elsevier

**#1 Elsevier’s Mendeley Data platform 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.

**#2 Researchers and institutions own and control all the data**

Mendeley Data allows you to keep data private, or publish it under one of **16 open data licenses**, so they stay in full control.

**#3 Mendeley Data is an open system**

It is a **flexible platform** — modules are designed to be used together, standalone, or combined with other RDM solutions.
#4 Mendeley Data can increase the exposure and impact of research

Mendeley Data Search indexes over **20 million datasets** from **1000+ repositories**

#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**

Want to know more?
elsevier.com/solutions/mendeley-data-platform/five-facts
Wrap up

- The value of Research Data Management (RDM)
- Mendeley Data for Institutions
- Mendeley Data and the FAIR principles
Remember to register:

Mendeley Data deep dive (5th November)

Questions?
Thank you

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