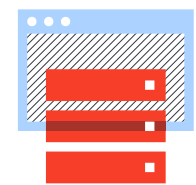


# Making your data FAIR with Mendeley Data

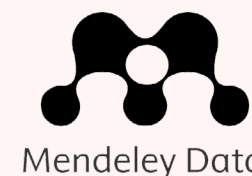


Librarians are essential to connecting researchers to new data, as well as expanding the reach of their institution's research. With a growing wealth of research data available, effective research data management (RDM) tools need to ensure datasets are easy to discover, share, and reuse by the research community.



## 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. The FAIR Data Principles were published in 2016, and now represent the gold standard for data sharing and re-use.



## 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

F

### Making your data **findable**

- Mendeley Data 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

A

### Making your 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.

I

### Making your data **interoperable**

#### Mendeley Data Repository:

- 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

R

### Making your data **reusable**

#### Mendeley Data Repository:

- 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



ELSEVIER

To view a full breakdown of how the Mendeley Data platform supports each of the FAIR data principles – please visit [www.elsevier.com/mendeleydatafairanalysis](http://www.elsevier.com/mendeleydatafairanalysis)

Downloaded from  
[www.elsevier.com/mendeleydatafairinfographic](http://www.elsevier.com/mendeleydatafairinfographic)

Elsevier and Mendeley Data are registered trademarks of Elsevier B.V. RELX Group and the RE symbol are trademarks of RELX Intellectual Properties SA, used under license. © 2020 Elsevier B.V. CC BY

Mendeley Data

# Supporting FAIR Data Principles in practice

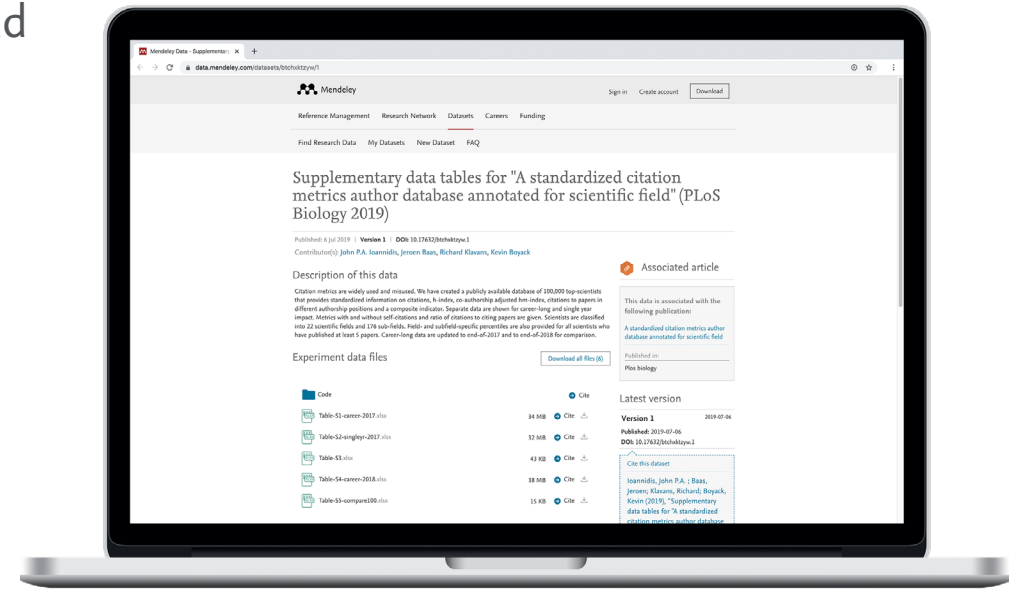
- F Findable
- A Accessible
- I Interoperable
- R Reusable

## Case Study Standardizing author citation metrics

[dx.doi.org/10.17632/btchxktzyw.1](https://dx.doi.org/10.17632/btchxktzyw.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:

- F A Dataset viewed over 100,000 times with 37,000 download
- F I Uses clear metadata descriptors and labeling
  - R Includes detailed reproducibility guidelines
- I R Folders used to separate code from data



*It was very straightforward to upload our datasets to Mendeley Data Repository – we didn't have to worry about anything, it just worked, even with the very large files that we had. One of the features that I appreciated the most was the ability to set a date for when the dataset became publicly available to everyone – while we were still confirming our final checks for publication, we could still share the data with external reviewers. The number of emails that I have received about these datasets is astronomical. A lot of people are excited about what we have done, and they are using the data for personal assessments or even institutional and higher-level assessments.*

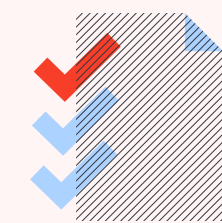
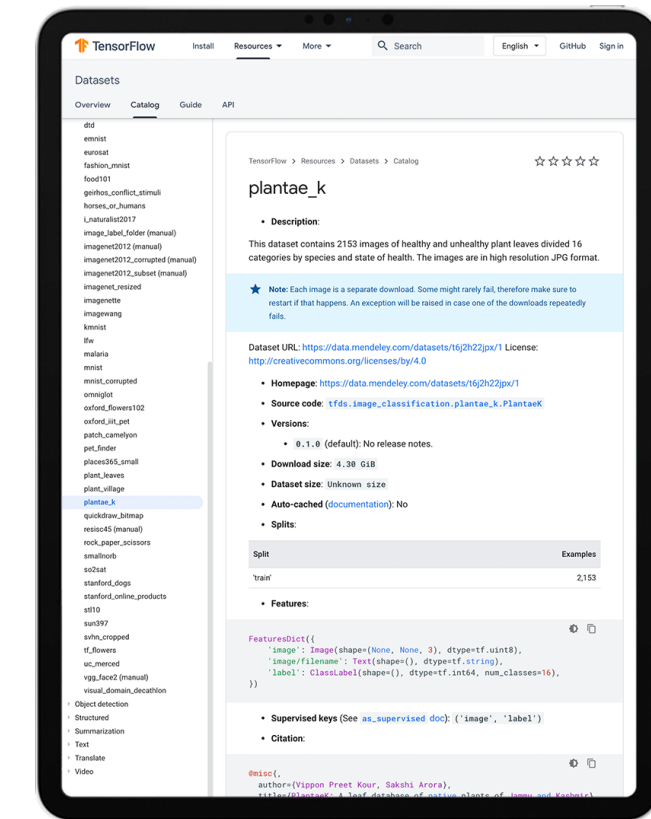
- John Ioannidis, Professor of Medicine, Health Research and Policy, Biomedical Data Science, and Statistics at Stanford University

## Case Study Reusing datasets to train deep-learning algorithms

[dx.doi.org/10.17632/hb74ynkjc.1](https://dx.doi.org/10.17632/hb74ynkjc.1)

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 this *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:

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



### Find out more

Visit [www.elsevier.com/FAIRwithMendeleyData](https://www.elsevier.com/FAIRwithMendeleyData)



To view a full breakdown of how the Mendeley Data platform supports each of the FAIR data principles – please visit [www.elsevier.com/mendeleydatafairanalysis](https://www.elsevier.com/mendeleydatafairanalysis)

Downloaded from [www.elsevier.com/mendeleydatafairinfographic](https://www.elsevier.com/mendeleydatafairinfographic)

Elsevier and Mendeley Data are registered trademarks of Elsevier B.V. RELX Group and the RE symbol are trademarks of RELX Intellectual Properties SA, used under license. © 2020 Elsevier B.V. CC BY

Mendeley Data