RESEARCH DATA MANAGEMENT SERVICE GROUP
Comprehensive Data Management Planning & Services

Cornell University (http://www.cornell.edu) 🔍

Search

	🔍

✣

Preparing FAIR data for reuse and reproducibility

Research funders are increasingly encouraging sharing of data associated with research that is machine readable and consistent with the FAIR Data Principles (for example, the Final NIH Policy for Data Management and Sharing (https://grants.nih.gov/grants/guide/notice-files/NOT-OD-21-013.html) released in 2020). The FAIR Principles (https://www.go-fair.org/fair-principles/) describe how data can be organized and documented so they are more Findable, Accessible, Interoperable, and Reusable in logical ways by other users and computer systems. Preparing and sharing your data in line with the FAIR Data Principles can facilitate discovery and reuse of your research.
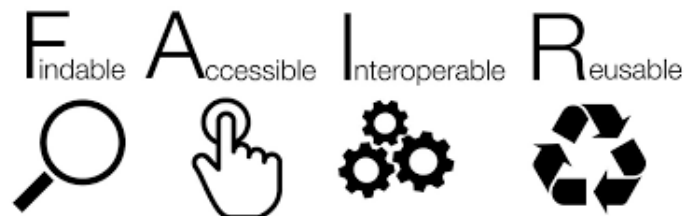
- Terminology
- How to make your data FAIR
- FAIR checklist
- References
- Related information

Terminology

FAIR Principles: the 'FAIR Guiding Principles for scientific data management and stewardship' (http://www.nature.com/articles/sdata201618) are a set of technical attributes published in *Scientific Data i*n 2016 to increase the Findability, Accessibility, Interoperability, and Reusability of data, emphasizing machine actionability due to our increasing reliance on computational systems when dealing with data.

- **Findable:** data and metadata are online and openly searchable with a persistent link that is uniquely attached to each specific dataset.
- **Accessible:** data and metadata are retrievable in machine-actionable form, with downloading options clearly described (including any needed authentication).
- **Interoperable:** data and metadata are consistently structured and described, both syntactically and semantically, so that algorithms can parse and ensure like data are accurately compared to like.
- **Reusable:** data and metadata are sufficiently annotated so machine and human users can determine fit-for-purpose in the context of their analysis.

Machine actionable: structuring data and content to make it possible for computational systems to find, access, interoperate, and reuse data without significant human intervention

Data interoperability: the capacity to which data can be analyzed and/or merged with similar data. Data interoperability relies on data standards, data documentation, and metadata to indicate to researchers which data sets or variables are comparable. (NNLM data thesaurus (https://old.nnlm.gov/data/thesaurus))

How to make your data FAIR

Technologies and supporting services for managing and sharing data, such as data repositories and collection software, are evolving rapidly. Depending on your area of research, you may wish to further explore and advance digital science methods and techniques in your own research. Regardless of your preferred tools and workflow, as the original researcher generating data that you are sharing, there are some key criteria to consider to ensure critical information that only you know is captured. It may be helpful to consider preparing your data in parallel with preparing your article manuscript as illustrated in the figure below.
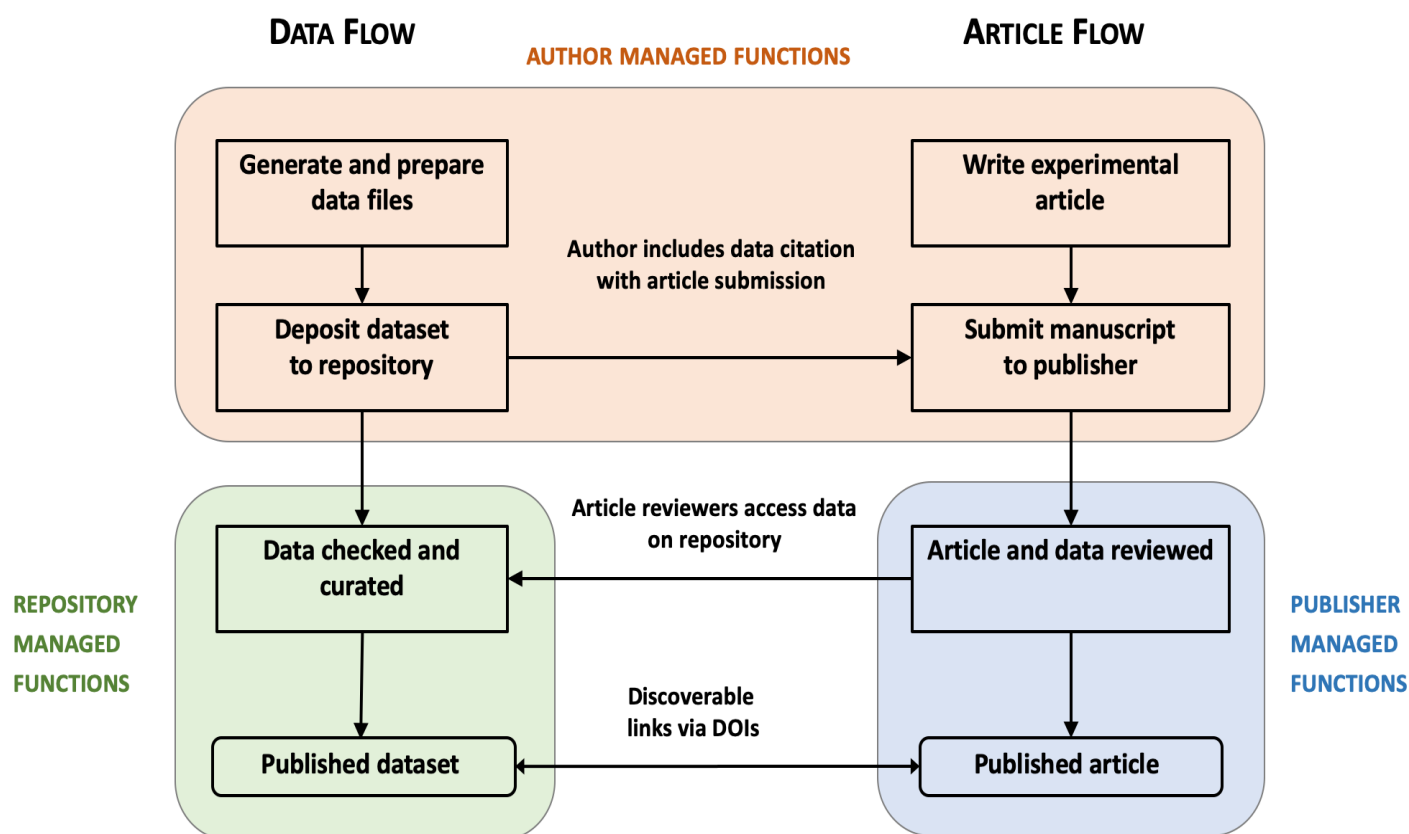
Preparing FAIR data starts long before you begin working on the final publication. Applying the FAIR Data Principles starts from good data management and documentation practices used throughout your research. A quick and basic checklist is provided below to see if your data files and documentation (i.e., metadata) support the FAIR Data Principles, followed by additional tips on how to prepare your data accordingly. More specific direction or requirements may be suggested by data repositories, certain journals, or within different disciplines. RDMSG consultants are available to help you navigate the process.

FAIR Checklist

Dataset/files ▶

☐ Is the dataset in an open & trusted repository (if available)?

☐ Does the dataset have a registered DOI?

☐ Are data files in standard and/or commonly available open formats (/content/file-formats) (as much as possible)?

☐ Are the data and/or metadata retrievable via an API and/or discoverable through an open search protocol (e.g., through Google)?

README/metadata ►

☐ Are all associated data files unambiguously named in the metadata and described including file types, software requirements and/or conversion information?

☐ Does the metadata include useful disciplinary notation and terminology? (e.g., SI units, common domain identifiers, explain acronyms, define field-specific jargon)

☐ Does the metadata include machine-readable standards where available (e.g. ORCIDs (https://guides.library.cornell.edu/orcid) for authors and/or data contributors, W3C/ISO 861 date standard (https://www.iso.org/iso-8601-date-and-time-format.html), and ITIS taxonomic IDs (https://www.itis.gov/))

☐ Are related articles referenced and linked in the metadata?

☐ Is a citation format (/content/data-citation) for the dataset provided?

☐ Are any license terms, attribution, or terms of use clearly indicated?

☐ Is the metadata exportable in a machine-readable structured text-based format? (e.g., XML, JSON)

☐ Dig deeper into more of our guidance about Metadata and describing data (/content/writing-metadata) or our Guide to writing "readme" style metadata (/content/readme)

Additional tips for preparing your data for sharing ►

Preparing your data files:

☐ You may choose to include raw data (as originally collected), processed data (e.g., signals encoded), or both. The decision depends on what is most useful or common in a discipline or specifically required by a publisher or repository

☐ Use file formats that are common and open (https://guides.library.cornell.edu/ecommons/formats) as much as possible, including for discipline-specific data types if open formats are available

☐ Use unambiguous filenames and organize the files logically according to your project (e.g., by sample, treatment, method, etc.); dig deeper in the file management (/content/file-management) guidance

☐ Explore more information about Preparing tabular data for description and archiving (/content/tabular-data)

Documenting your data and files:

☐ For an easy, low-barrier approach, use a ReadME template and save as a plain text document (.txt). Note that some repositories may provide specific documentation templates.

☐ List the data files included in the package, and/or describe the file naming schema and organization. Include their formats and any specific software requirements and/or conversion information if you have it.

☐ Describe methods of data collection and file structures and organization including useful notation about the data headers, units, sample identifiers, etc. Use standard or conventional terminology or nomenclature in your discipline.

☐ Reference associated articles, code and related datasets. Include ORCIDs of all data contributors.

☐ Find much more in our Guide to writing "readme" style metadata (/content/readme)

Depositing your data in a repository:

☐ Select a reputable data repository and upload dataset (publishers or funders may require specific repositories; many domain-specific repositories provide enhanced services and curation for specific data types)

☐ Make sure the repository provides a persistent identifier (e.g., DOI, handle, or other) and specifies conditions for others to access and re-use the data (such as a public domain declaration or Creative Commons attribution license; licensing policy may vary by repository)

☐ Provide a pre-formatted citation (/content/data-citation) (and license attribution (/content/intellectual-property) if appropriate) for the dataset on your website and other materials so users can easily copy and attribute; for example:

- Author(s). Dataset Title, Version. Data Repository (or Journal if appropriate). Year. DOI. (Date accessed)
- [License attribution if appropriate.]

☐ Learn more about Sharing and archiving data (/content/sharing-and-archiving-data)

References

Wilkinson, M., Dumontier, M., Aalbersberg, I. *et al.* The FAIR Guiding Principles for scientific data management and stewardship (https://doi.org/10.1038/sdata.2016.18). *Sci Data* 3, 160018 (2016).

FAIR Principles (https://www.go-fair.org/fair-principles/). GO FAIR.

Related information

- Data citation (/content/data-citation) (Cornell Research Data Management Service Group): Examples of different data citation formats.
- Data storage and backup (/content/data-storage-and-backup) (Cornell Research Data Management Service Group): Things to consider when planning for storage and backup.
- File formats for preservation (/content/file-formats-preservation) (Cornell Research Data Management Service Group): Guidelines for selecting file formats for preservation and reuse.
- File management (/content/file-management) (Cornell Research Data Management Service Group): Guidance on file organization and naming conventions.
- Frequently asked questions (/content/frequently-asked-questions) (Cornell Research Data Management Service Group): Addresses more questions about data sharing.
- Guide to writing "readme" style metadata (/content/readme) (Cornell Research Data Management Service Group): Recommended minimum content to include and a downloadable template.
- Introduction to intellectual property rights in data management (/content/intellectual-property) (Cornell Research Data Management Service Group): Data licensing to facilitate reuse.
- Metadata and describing data (/content/writing-metadata) (Cornell Research Data Management Service Group): Information about documenting data for future reusability.
- Preparing tabular data for description and archiving (/content/tabular-data) (Cornell Research Data Management Service Group): How to maximize the likelihood of long-term preservation and potential for reuse.
- Sharing and archiving data (/content/sharing-and-archiving-data) (Cornell Research Data Management Service Group): Includes information on choosing a repository and documenting conditions for reuse.

Page last updated Oct. 2022.

(http://www.twitter.com/curdmsg)

- Home (/) /
- Contact RDMSG (https://data.research.cornell.edu/content/contact-rdmsg) /
- Join our mailing list (/content/mailing-list) /
- Privacy Statement (/content/privacy) /
- FAQ (/content/frequently-asked-questions) /
- Site Map (/content/site-index) /
- Web Accessibility Assistance (https://www.library.cornell.edu/web-accessibility)