

Documenting & Sharing Accessible Data and Visualizations

A brief guide on best practices, tips, and resources for documenting and sharing datasets and visualizations

Created for the RDAP workshop “Accessibly Create and FAIR(Iy) Share Visualizations” delivered on January 23, 2025.

Introduction

This guide covers in more detail the concepts of data viz documentation outlined in the RDAP workshop “Accessibly Create and FAIR(Iy) Share Visualizations” delivered on January 23, 2025. It includes overviews of concepts, documentation examples, and resources for further reading.

Creating an Accessible Ecosystem for Your Viz

It is important to consider not just your visualization, but all supporting data and attendant documentation. Consider the following:

- Alone, your visualization cannot be truly accessible.
- Link to the data behind your visualization and ensure it is accessible.
- Accessibility cannot exist without documentation.

To make it FAIR, you want to make these documents and data findable, accessible, interoperable, and reusable.

Documentation

Documentation is an essential step in ensuring that any research data is findable, accessible, reusable, reproducible, and enduring. This is especially true of data visualizations, which are easily orphaned from their sources. The level of documentation necessary varies from project to project, but even a minimal amount can make a difference in the overall usability of a visualization. When presenting a visualization, or series of visualizations, it is wise to at least answer the following questions:

- What is the source of the visualization?
- Who is responsible for the source data?
- How can someone access and use that data?

Documentation can include READMEs, data dictionaries, codebooks, taxonomies, metadata schemes, field notes, and any other documents that help users interpret, understand, and reuse your data effectively.

Types of Documentation: The Basics

- **README:** document detailing the provenance of your data, what software are needed to use it, and any other information necessary to understand the project. The complexity of this type of document can range from a simple abstract to a document detailing everything from access restrictions to file hierarchy logic. This is best written as a simple text document (.txt) or a markdown (.md) file if formatting is required.
- **Data dictionary:** a tabular document describing the variables and table elements in a dataset using standardized language. This is best saved as a comma separated value file (.csv), following UTF-8 encoding, without distinct worksheet tabs, or as a .txt file within a README.
- **Codebook:** similar to a data dictionary but with more human-readable information that goes in to greater detail about all used data elements, provenance, collection methodologies, and field notes. Think of it as a *super* README.

Examples of Documentation

- [Carnegie Mellon University: Example README](#) which includes elements like:
 - Author contact information
 - File naming convention
 - Descriptions of each file
 - Variables
 - Methodology
 - Reproducibility notes
- [README template from the University of Rochester](#) which includes elements like:
 - Funder information
 - File types
 - Required software
 - Folder hierarchy logic
- [Data dictionary example from Nature](#) which includes elements like:
 - Variable names
 - Variable definitions
- [Codebook example from USGS](#) which includes elements like:
 - Longform definitions of all data elements
 - Delineations of element types
 - Summaries

Note: data dictionaries and codebooks are terms often used interchangeably, but the nomenclature is less important than ensuring these data and metadata are included *somewhere* in a logical straightforward manner.

Interoperability

Interoperability is the characteristic of a program, file, or set of files allowing a wide array of users to render and manipulate them over time. This includes opensource software, file formats unencumbered by licensing restrictions, or file types that have existed – and are *most likely* – to exist over a long period of time. Beyond the actual files, it's important to understand that interoperability extends to the visual logic of things like file naming conventions and folder hierarchies. When considering interoperability it is wise to consider the following:

- Can files be rendered or opened by widely available software?
- Do files comply with recognized standards within a discipline?
- Are files organized and labeled so that a total stranger can correctly glean their logic?
- Are the files themselves created in such a way that assistive technology can read and interpret them?

A good rule of thumb when exporting data is to use the simplest, most widely accessible format so long as it doesn't affect the quality and usability of the data. For example: converting some DICOM-compliant files like .vsi, or even less complex files like .prism for graphing, to static images like .tiffs would not make them more accessible, as those types of files are dynamic and cannot be effectively rendered outside of their original environment.

Below is a table detailing generally preferred file formats and their required software. This is by no means exhaustive, covers much of what may appear in the average dataset.

File Formats

File Type	Preferred Format	Software	More Info
Word Processing	PDF/A (.pdf); Open Document Format (.odf); MS Word Open XML (.docx)	Adobe Acrobat Reader, MS Word, Apache OpenOffice	Word processing documents are any text-based works that rely on formatting. The listed formats are widely accepted and versatile. Avoid embedding audio or video files in PDFs.
Text	Plain UTF-8 text (.txt); plain text with markdown formatting	Notepad, MS Word, Apache Open Office	Plain text is the most widely accepted way to format and distribute textual data, with or without formatting notation.
Data & Datasets	Delimited flat file (.csv, .txt)	MS Excel, Notepad	This is specific to tabular data and includes tab- and comma-separated formats. Some datasets cannot be represented so simply, so maintain preference for whatever best serves the data (e.g., ACCDB, SAS).

Raster & Vector Images	PNG (.png); TIFF (.tif); JPEG2000 (.jp2); Scaled Vector Graphics (.svg)	Adobe Photoshop and/or Illustrator, MS Paint, IrfanView	For images, whatever format in which the work was originally exported but prefer uncompressed losslessly compressed formats like those listed.
Audio	WAVE (.wav); AIFF (.aiff); FLAC (.flac)	VLC, Windows Media Play	For audio files, whatever format in which the work was originally exported but prefer uncompressed or losslessly compressed formats like those listed.
Video	AVI (.avi); MKV (.mkv), Motion JPEG2000 (.mj2), MP4 (.mp4, .m4v)	VLC, Windows Media Player, Adobe Premiere	For video files, whatever format in which the work was originally exported but prefer uncompressed or losslessly compressed formats like those listed (*MP4 video is compressed, but universally accepted).
Presentations/ Slide Decks	MS PowerPoint Open XML (.pptx); PDF/A (.pdf)	Adobe Acrobat Reader, MS PowerPoint	Like word processing documents, prefer the most accessible versions that preserve formatting and are faithful to their original presentation.
Spreadsheets	CSV (.csv); MS Excel Open XML (.xlsx)	MS Excel, Notepad	Like tabular datasets, spreadsheets are best preserved as comma-separated value files or MS Excel Open XML files without multiple worksheets.

Organization

For smaller datasets compiled with minimal use of proprietary or specialized software, folder organization can easily be rendered by sticking to a folder and file naming convention. Doing so early on the collection and analysis process will save time and reinforce adherence to the process in subsequent efforts. Often these naming conventions nullify the need for explanation in a README, especially if they are self-evident.

For example, a folder labeled “**202501023_Core10DataProject_TabularData**” was is dated, has a project name associated with it, and expresses the contents. Furthermore, a file within that folder may read

“202501023_Core10DataProject_Dataset_v001.csv” which delivers most of the same information with the addition of what it is, a version declaration, and clear file suffix.

However, it is common for some data analysis and collection software to export folders and files with names that are perfectly readable to that program, but inscrutable to humans. In the case of larger datasets with hundreds of folders and subfolders, and thousands of files, it is unreasonable to attempt to rename and reorganize the data. Not only is such an effort time-inefficient, but could disrupt the stability of the data when rendered in the necessary software. In this case, it is advisable to explain the hierarchy and folder contents (on average) in a README to best of your ability, or indicate that the required software will correctly interpret what is otherwise an enigma.

Accessibility

Accessibility refers to a file’s nature with respect to the many assistive technologies used by persons with disabilities. It also refers to the long-term technical accessibility of that file. Making your dataset and attendant documentation accessible in all forms will broaden access and help ensure long-term stability. At a file-level, you can accomplish accessibility in all its forms with features and tools like:

- Alternative text (alt-text)
- Optical character recognition
- Color contrast checkers
- Color blindness checkers
- Close captions
- Subtitles
- Descriptive audio

In the case of simple textual documentation like docs, PDFs, or spreadsheets, accessibility checkers are built in to common tools like Word, Excel, and Acrobat Pro. Often, following simple habits like utilizing standardized headings, avoiding low contrast colors, and preemptively adding alt-text will make accessibility checkers superfluous as you create your documentation.

A more in-depth walkthrough of Adobe Acrobat Pro’s accessibility checker can be found in the accompanying guide “Creating Accessible PDFs Using Acrobat Pro: Demo Walkthrough” labeled 20250123_RDAP_PDFguide.pdf.

Sharing Your Work

For your data reap the benefits of being accessible and well-documented, it must also be shared. Depending on the nature of your data, it may only be allowed to be shared amongst a specific group of users, but there are many extant repositories that can accommodate a range of usage requirements. When considering where to share it is wise to look for repositories that offer some or all of the following:

- DOIs
- Format agnosticism
- Retention periods
- Search engine indexing

- Funder endorsements
- Curation support
- Licensing options
- Cost
- Disciplinary focus
- And more...

The best repository will exist at the nexus of what is required by your funder, how the data can be legally *and* ethically shared, and what best serves the needs of those data.

Further Reading

Below is a list of resources and tools for assisting you in ensuring you are following best practices in data documentation, accessibility, and sharing.

Documentation

- [Cornell University guide to writing README files](#)
- [Dataset README template from GitHub user shashvatshah9](#)
- [OSF data dictionary guidance](#)
- [McGill University Codebook Cookbook](#)

Interoperability and Organization

- [Library of Congress Sustainability of Digital Formats](#)
- [Library of Congress Recommended Formats Statement](#)
- [University of Rochester LibGuide on Organizing Data](#)
- [COPTR's file format identification tools](#)

Accessibility

- [WebAIM Contract Checker](#)
- [Coblis Color Blindness Simulator](#)
- [Adobe Acrobat Create and Verify PDF Accessibility](#)
- [MS Office Suite accessibility checker](#)

Repositories

- [NIH NLM Open Repository Database](#)
- [University of Rochester Data Sharing & Storage LibGuide](#)

Created by [Matthew C. Mariner](#), Data Curator Librarian, University of Rochester Libraries

Created under a [Creative Commons CC-BY 4.0 License](#)

Last updated 2025/01