Why do open formats matter?

Research data should be made available in open formats that can be easily read, written, parsed, and displayed by a computer. Data management policies must conform to standards that manage the risk of file format obsolescence or degradation. Therefore, choosing a suitable file format for data preservation and sharing is vital for access and reuse.

What is an open format?

A file format with a freely available published specification that places no restrictions, monetary or otherwise, upon its use or, at the very least, can be processed with at least one free/libre/open-source software tool.

Lossless vs. lossy formats

A lossless format retains the original detail of the data file (e.g., TIFF and WAV).
A lossy format discards information permanently to lower the quality of that data and ease sharing and visualization (e.g., JPG and MP3).

Best practices for file format selection:

- Choose non-proprietary, open, and documented standards
- Prefer unencrypted and uncompressed formats
- Avoid embedded files, programs, or scripts
- Use standard character encodings – ASCII, Unicode, UTF-8

Tips for handling file formats:

- Retain your original unedited raw data in its native formats as your source data
- Document the tools, instruments, or software used for creating files
- Create open equivalents of files in proprietary and closed formats

Maximize the longevity and interoperability of your data using open formats!

For a more comprehensive list of open formats, visit: tiny.cc/open-file-formats

What if a format becomes obsolete?

Try using the old computer hardware along with the operating system and all the required software, or use an emulation system to recreate the native environment.

Want to learn more? Contact us: rds@library.ucsb.edu