



Digital Archiving and Preservation

File Format Longevity

Careful consideration of file formats is important, whether you have years of digital data or are about to embark on a new project. Working with digital material over the years can result in files saved in a number of formats, some of which may be proprietary to a single program. Yet, even common programs such as Word or PDFs can change their underlying format, rendering older files difficult or impossible to access. Below are some guidelines to assist you in selecting appropriate formats for a variety of digital data, including documents, images, audio files, and general data.

Documents

- **Microsoft Word** often is program of choice for text documents and should be supported into the foreseeable future. Nevertheless, each time you upgrade Word / Office (e.g., from 2007 to 2010), you should open your important files and re-save them to ensure that they remain viable. This is especially important if a new format is introduced (e.g., .docx instead of .doc).
- Having copies of important documents in **PDF** format is also useful, as that is an ubiquitous format that can be opened by a variety of programs.

Images

- **JPG** or **PNG** files are standard formats that should be viable for decades at least.
- **TIFF** files provide a “lossless” (i.e., uncompressed) image that makes a great archival “master,” and is most useful if coming from a digital camera’s RAW setting or a personal scanner.

Audio

- **MP3** files are a widely-adopted standard that most common programs (e.g., Audacity, VLC) can save as.
- **WAV** files are a “lossless” (i.e., uncompressed) format that makes a great archival “master.”

General Data / Spreadsheets

- Tabular data are often saved in **Excel**, which certainly will be supported into the future. However, Excel often makes assumptions about cell formatting (e.g., numeric, currency, text) that can affect sharing your data or importing it into other programs (e.g., R, SPSS, ArcGIS).
- As such, having a copy of your master data in a **plain-text delimited file** (e.g., comma or tab-delimited) can be helpful. These are generally saved with a .TXT or .CSV extension, and Excel can do this via Save as...