

## Making Digital Information Last: Good Practices for Information Management

Digital information is more fragile than information produced on paper and most other tangible formats. Digital storage media such as CDs and DVDs have shorter life spans, and accessing digital information requires software and hardware that may no longer be available when the file is needed.

However, the choices that we make when we create and store documents, image files, audio and video, and datasets can improve their likelihood of remaining usable (and findable) over long periods of time.

Most software includes settings that can affect the long-term usability of the files it creates. *Please consult with your ITS, CTS, or Library liaison for further information or advice.*

**1. Appraisal:** Is the file that you're creating one that you or someone else might want to use two years, five years, or more in the future?

✓ Consider future uses for working papers, research datasets, or pre-publication versions of articles, books, and reports.

**2. Select Well-Supported Formats:** Some file formats are more likely to remain usable over long periods than others. In general, proprietary formats -- formats that are owned by companies -- are less likely to remain usable than non-proprietary formats, and "closed" formats -- whose specifications are corporate secrets -- are less likely to remain usable than open formats whose specifications are public.

Preferred formats for master files with a **high likelihood of remaining usable:**

- Text: PDF (but see notes below), Postscript, XML, ASCII (.txt), Rich Text (.rtf)
- Data: delimited ASCII, SGML, XML
- Databases: Databases are unique and pose special technical challenges for both preservation and future migration that require special consultation.

Common examples of formats with **less likelihood of remaining usable:**

- Text: MS Word (.doc), Word Perfect (.wpd)
- Data: Excel (.xls)
- Image: Photoshop
- Other Media: Powerpoint (.ppt)

**3. PDF Files** require a few simple steps to improve the likelihood that they will remain usable into the future. Minimize the number of external applications or other information that the file needs to open, and select those with the greatest likelihood of being available to the user (see above). Make sure that your settings include:

- Minimum 600 dpi resolution
- Print optimized
- Embedded fonts: All fonts used must be embedded in the document. This means that all of the font information used to make your document look the way it does is stored in the PDF file (otherwise, it may not be available on the computer being used by someone who wants to read it).

**4. Media Files (images, movies, and audio)** present special challenges. The most information-rich formats – uncompressed -- typically create files too large to deliver on course pages or via the Web.

**Consider creating “master” versions of important files in uncompressed formats, from which you derive “deliverable” content in smaller compressed versions. Store the uncompressed versions offline, and use the smaller versions for Web pages or Pioneer Web.**

Examples:

<i>Still Images</i>	<i>Video</i>	<i>Audio</i>
Master Version: TIFF (uncompressed)	Master Version: digital video (DV) format or the original or digitized format with its native resolution and frame rate.	Master Version: WAV, Apple lossless, or lossless AAC
Service or Deliverable Version: JPEG	Service or Deliverable Version: MP4 file with MPEG-4 video and AAC audio	Service or Deliverable Version: MP3

**5. Describe What You Create:** Providing basic information about your files – individually or in batches -- will increase the chances that they can continue to be used over time. Try to include the name of the creator, a description of the content, the date created, the last date modified, the name of the file, and the format of the file.

- For files stored on a server, create filenames and file folder titles that will be meaningful to users who come after you.
- Create “Readme” files in .txt (ASCII) format to document what you have done.
- For formatted data (like datasets), create data dictionaries that describe the data fields and expected contents -- paper or online.
- For files stored on portable media (disks, DVDs, tape) attach a paper label and print out a directory.

**6. Select Secure Storage:** In general, files stored on networked servers maintained by professional administrators adhering to best practices are more likely to remain accessible than files held on isolated disks (including the hard drive of your computer).

- If you must store important files on a workstation hard drive, back up regularly onto high quality removable media, such as CD-ROM, DVD, or DLT tape. Consider making redundant backup copies, perhaps in more than one format, such as DVD and DLT, and storing those in different buildings.

**7. Want to Learn More?** Start with the online tutorial *Digital Preservation Management: Implementing Short-term Strategies for Long-term Problems* (Cornell University Library) at [http://www.library.cornell.edu/iris/tutorial/dpm/eng\\_index.html](http://www.library.cornell.edu/iris/tutorial/dpm/eng_index.html)