Criteria for Preservation File Formats

1. Independent

Technological independence is important in a preservation file format because that means the format will be usable in a variety of technological environments now and possibly into the future. If a file format can be viewed and manipulated only with certain hardware or software, it will be more difficult to manage over time. Also, the greater the degree to which a file format depends on external data sources to represent itself correctly, the less valuable it will be as a preservation format. For example, the PDF/A file format stores all necessary fonts within the body of each file, so that it can be presented properly even if not being viewed on a computer already storing those fonts.

2. Long-lived

To some degree, the longer a file format has been in existence, the better it is as a preservation format—all other considerations being equal. The longer a format has been around, the more likely it is that there is good support for the format and that it will be widely used. However, this value goes only so far, since all things digital eventually change or disappear. For instance, the TIFF image file format has existed for decades, making it unlikely that it will disappear soon, but TIFF images are rarely used in standard personal or institutional settings anymore.

3. Metadata-supporting

The ability of a file format to store metadata within itself is a measure of the value of the format of a preservation format. A self-documenting digital file is one that always creates metadata to identify the file as the format it is and to identify the basic metadata about its contents. For instance, a PDF/A always stores within its metadata information to distinguish it from other types of PDFs as well as other required metadata, such as date of creation.

4. Open

The more open a format is, the more likely it is that you will be able to reverse engineer a way to see a file in that format if need be in the future. Open formats are often international standards of one kind or another, which sometimes also means they are fairly ubiquitous. In the end, what open means is that full access is available to the documentation for the format along with the unrestricted legal ability to use that documentation to develop means to both see and create a file in that format.

5. Transparent

The easier it is for someone to review the inner workings of a digital file, the more likely it is a good preservation format. Uncompressed and unencrypted files are more transparent, because they do not provide additional barriers to access and it is always best for a preservation copy of a digital file to be saved in an uncompressed and unencrypted form.

6. Ubiquitous

The more widely adopted a format is, the more likely it is that technical support for the format will be available from Information Technology Staff or outside vendors. For this reason, new formats provide challenges for preservation.

Common Digital File Formats and Preservation Formats

The charts below are organized by the data format of a digital record (audio, text, still image, etc.). Review for each data format the common file formats likely to be used in your institutions, review acceptable storage formats you can use if you find it difficult to use a preservation format, and then provide a list of preferred preservation formats. Note that the preservation formats recommended in these tables are ones that comply with all or virtually all of the criteria for digital formats listed in Appendix 1 above. The chart also lists additional considerations that you must examine for specific data formats.

Data Format	Common File Formats	Acceptable Storage Formats	Preferred Preservation Formats
AUDIO	MP3 (.mp3), Waveform Audio File Format (.wav), Windows Media Audio (.wma)	Windows Media Audio (.wma)	Audio Interchange File Format (.aif, .aiff), Waveform Audio File Format (.wav)
Additional Considerations for Audio Ability to support high audio resolution Ability to support multi-channel recording			

Data Format	Common File Formats	Acceptable Storage Formats	Preferred Preservation Formats
DATABASES	dBase File (.dbf), Microsoft Access (.mdb), OOXML Access (.mdbx)	OpenDocument Format (.odb), OOXML Access (.mdbx)	Comma-separated Values File (.csv), Tab-separated Values File (.tsv, .tab), Structured Query Language (.sql)
Additional Considerations for Databases Ability to support relationships within the file (between tables and between fields) Ability to preserve the ability to manipulate the data			

Data Format	Common File Formats	Acceptable Storage Formats	Preferred Preservation Formats
SPREADSHEETS	Excel (.xls), OOXML Excel (.xlsx)	OpenDocument Format (.ods, .fods), OOXML Excel (.xlsx)	OpenDocument Format (.ods, .fods), OOXML Excel (.xlsx)
Additional Considerations for Spreadsheets Ability to document and store mathematical formulae Ability to support relationships within the file			

Data Format	Common File Formats	Acceptable Storage Formats	Preferred Preservation

			Format
STILL IMAGE	Bitmap Image File (.bmp), Graphics Interchange Format (.gif), JPEG (.jpg, .jp2), Portable Network Graphics (.png), Raw Image Format (many proprietary extensions) TIFF (.tif)	JPEG (.jpg,), JPEG 2000 (.jpx, .jp2), PDF/A (.pdf), PNG (.png)	TIFF (.tif)
Additional Considerations for Still Images Ability to support high image resolution Ability to record a full range of color			

Data Format	**Common File Formats**	**Acceptable Storage Formats**	**Preferred Preservation Formats**
TEXT	ASCII (.txt), OOXML Word (.docx), PDF plain (.pdf) , Rich Text Format (.rtf), Word (.doc)	Open Document Format (.odt, .fodt), PDF plain, with fonts embedded (.pdf)	ASCII (.txt), Extensible Markup Language (.xml), PDF/A-1 (.pdf), PDF/A-2 (.pdf), PDF/A-3 (.pdf), Unicode
Additional Considerations for Text The relative importance of preserving the overall look and feel of the document Ability of the format to preserve fonts, color, tables, and other document features Ability to save the navigation and structure (headings, chapters, etc.) of the document			

Data Format	**Common File Formats**	**Acceptable Storage Formats**	**Preferred Preservation Formats**
VIDEO	Audio Video Interleaved (.avi), QuickTime Movie (.mov), MPEG-4 Part 14 (.mp4), Real Video (.rv) Windows Media Video (.wmv)	Audio Video Interleaved (.avi), MPEG-1, MPEG-2 (.mpg, .mpeg)	Motion JPEG2000 (.mj2)
Additional Considerations for Video Ability to support high in terms of image and audio resolution Ability to record a full range of color			

Ability to support multi-channel recording