A Framework Enabling the Preservation of Electronic Records

Elizabeth England and Leslie Johnston
National Archives and Records Administration

Assessing Risk in the Digital Preservation Framework

- In 2018 NARA established a Holdings Profile to document the formats in its holdings.
- A Risk and Prioritization Matrix is used to apply a series of weighted factors related to the preservation sustainability of a file.
- Each factor is weighted using relative weightings that map to the level of risk for each question and, to the extent that it can be defined, cost. The Matrix is designed to analyze preservation risks:
  - How much publicly available documentation is there for a format?
  - What impacts NARA's ability to process, render, and preserve that format?
  - How prevalent is the format in NARA's holdings, and what capabilities and capacity exist in-house to process and preserve it?
- A numeric score is generated, mapped to High, Moderate, and Low Risk.

NARA Guidance: Preferred	NARA Guidance: Acceptable	Numeric Risk Rating	Risk Level	NARA Format ID	Format Name	File Extension(s)	Category/Plan(s)	Is the format proprietary?	Does the format have a published open specification?	Are there available tools that can validate the technical integrity of a file encoded in this format against the published specification?	Has the specification been approved and published by an internationally recognized standards body?	Is the available specification complete and accurate?	Total Disclosure Score. Highest possible score = 10; Lowest possible score = -6
X		47.00	Low Risk	NF00540	Broadcast Wave (BWF) v. 2	wav	Digital Audio	-1	2	2	2	2	7
		-17.00	High Risk	NF00713	Broderbund The Print Shop Letterhead	let	Presentation and Publishing	-1	-1	-1	-2	0	-5
X		14.00	Moderate Risk	NF00636	BSB	kap; cap; bsb	Navigational Charts	2	2	0	2	2	8
		-1.00	Moderate Risk	NF00137	Btrieve 5.1	btr	Databases	-1	-1	2	-2	0	-2
		-1.00	Moderate Risk	NF00541	Btrieve 6.0	btr	Databases	-1	-1	2	-2	0	-2
		-1.00	Moderate Risk	NF00542	Btrieve 6.1	btr	Databases	-1	-1	2	-2	0	-2
		13.00	Moderate Risk	NF00732	C/C++/Objective-C Header File	h	Software and Code	2	-1	-1	-2	0	-2
		27.00	Low Risk	NF00138	C# Code	cs	Software and Code	-1	2	2	-2	2	3
		21.00	Moderate Risk	NF00139	CALS Compressed Bitmap	cal; ct1; ct2	Digital Still Image	2	2	2	0	2	8
		-22.00	High Risk	NF00468	Canon RAW 1.0	crw	Digital Still Image	-1	-1	-1	-2	-1	-6
		-16.00	High Risk	NF00469	Canon RAW 2.0	cr2	Digital Still Image	-1	-1	-1	-2	-1	-6
		25.00	Low Risk	NF00141	Cascading Style Sheets 1.0	css	Web Records; Software and Co	2	2	0	2	2	8

Preservation Action Plans in the Framework

The Plans document file format risk, and collate links to specifications and other digital preservation resources. The recommended preservation tools and actions for formats included in the Plans are based on current NARA decisions and capabilities. The Plans consist of two sets of documents:

- **Record Category Plans**: Documentation of Significant Properties for 16 record categories (Email, Databases, Still Images, etc) - Appearance, Structure, Behavior, and Context. These are the properties that should, if possible, be retained in any format migration, and are used as metrics to test potential tools for the preservation migrations.

- **Preservation Action Plans**: A single spreadsheet containing over 650 file formats across all record categories, containing the specifications, resource links, format information, and preservation actions for the formats to ensure the Framework's actionability and extensibility to other institutions.

Preservation Action Plan: Digital Still Image Records
National Archives and Records Administration (NARA)

Plan Date: 20210525
Template: 202105

Digital Still Image Records

Digital still images are digitally encoded representations of the tonal and brightness information of a subject into a bitmap. Data from digital cameras and scanning devices record light characteristics as numerical values into a grid or raster of picture elements (pixels). There are two types of raster file digital image categories: Digital still photographs of natural, real-world scenes or subjects produced by digital cameras, and scanned images of textual documents, illustrations, posters, graphics, cartographic records, photographic prints, slides, and negatives. Image file formats are standardized means of organizing and storing rasterized data that can be used on a computer display or printer.

The term raster data is often contrasted with vector data, in which geometrical points, lines, curves, and shapes are based upon mathematical equations, thus creating an image without specific mapping of data to pixel. Bit-depth, spatial resolution, and color encoding, for example, are all important properties of still images.

Significant Properties of Digital Still Image/Raster Images

To render an authentic digital photograph one must preserve the structural, technical, and descriptive metadata that allow certain appearance properties to persist. Many of the properties native to raster image file formats are the result of industry efforts to develop common standards and interoperability. Many file formats for digital photography and scanning are the same except that most digital cameras create native camera raw proprietary formats, JPEG, and DNG, and rarely TIFF, JPEG2000, PNG, or GIF.

Appearance is a critical property for this category due to the common purpose or use of this type of record is to depict scenic information or to render the informational and artifactual aspects of a scanned original. Tone fidelity, resolution, bit depth, color encoding as well as compression algorithms all contribute to the preservation of the file. There is widespread adoption of most formats and rendering and display platforms. A unique property for scanned multi-page documents is descriptive and administrative metadata that may be held external to the electronic record and could be a risk for long term identification of the context.

Appearance

Name	Definition	Function Description
Size	Determined by bit-depth, spatial resolution, compression, and color encoding.	
Color	Color mode, color space.	Mathematical representations of color information needed to encode and decode color information such as Hue, Chroma, lightness, white point.
Bit-depth	The number of bits used to indicate color and tone information of a pixel.	High or low bit depth contributes to the pleasing transformation of color accuracy, gradients, and tonal information. Also greatly affects issues such as signal clipping and transformative image editing.
Orientation	Portrait versus Landscape.	

Structure

Name	Definition	Function Description
Layout Structure	Embedded technical metadata captured at the time of creation describing, among other things: File format/encoding; Compression; Resolution; Bit depth; and EXIF (Exchangeable Image File Format) information.	

Format Name	Extension(s)	Category/Plan(s)	NARA Format ID	MIME type(s)	Specification/ Standard URL	PRONOM URL	LOC URL	British Library URL	WikiData URL	ArchiveTeam URL
Broadcast Wave (BWF) v. 2	wav	Digital Audio	NF00540	audio/x-wav	https://tech.ebu.ch/docs/tech/tech3285.pdf	https://www.nationalarchives.gov.uk/pronom/fmt/527	https://www.loc.gov/preservation/digital/formats/fdd/fdd000357.shtml		https://www.wikidata.org/wiki/Q27526504	http://fileformats.archiveteam.org/wiki/BWF
Broderbund The Print Shop Letterhead	let	Presentation and Publishing	NF00713			https://www.nationalarchives.gov.uk/pronom/fmt/1300				http://fileformats.archiveteam.org/wiki/The_Print_Shop
BSB	kap; cap; bsb	Navigational Charts	NF00636		http://libbsb.sourceforge.net/bsb_file_format.html				https://www.wikidata.org/wiki/Q27823992	http://fileformats.archiveteam.org/wiki/BSB
Btrieve 5.1	btr	Databases	NF00137	application/octet-stream		https://www.nationalarchives.gov.uk/pronom/x-fmt/308			https://www.wikidata.org/wiki/Q48805099	http://fileformats.archiveteam.org/wiki/Ext:btr

Format Name	Risk Level	Preservation Action	Proposed Preservation Plan	Description and Justification	Preferred Processing and Transformation Tool(s)
Broadcast Wave (BWF) v. 2	Low Risk	Retain	Retain	Broadcast Wave (BWF) all versions 0, 1, and 2 are considered preferred formats as per NARA Transfer Guidance. Format is ubiquitous and well-documented.Version 2 is a substantial revision of Version 1 which incorporates loudness metadata (in accordance with EBU R 128). This version is fully compatible with Versions 0 and 1, but users who wish to ensure that their files meet the requirements of EBU Recommendation R 128 will need to ensure that their systems can read and write the loudness metadata.	FFmpeg; VLC; Windows Media Player
Broderbund The Print Shop Letterhead	High Risk	Transform	Transform to a TBD format, possibly PDF	The Print Shop is a desktop publishing application from Broderbund. The .let file is a letterhead project file present from versions 10-23.1 of The Print Shop. All files from these versions are in the OLE2 document container format.	The Print Shop
BSB	Moderate Risk	Retain	Retain	BSB is a format used by some hydrographic offices in the production of Raster Nautical Charts (RNCs). Although primarily used for marine navigation, the format is capable of representing a wide variety of maps and charts. BSB consists of a documentation file and one or more image files. The binary raster files of chart information are .kap (or .cap), and ASCII .bsb files contain data about the raster file. The format was originally developed by MapTech, but it's now an open standard.	FME Desktop or Server
Btrieve 5.1	Moderate Risk	Transform	Transform to CSV	A BTR file is a proprietary database file created by the Btrieve, transactional database program BTR files generally have an associated set of DDF data definition files that describe the database structure. Btrieve 12 is still supported, but no new products/enhancement have been released since 2015.	Procure and/or develop tools. Any process will require Actian (formerly Pervasive) Btrieve drivers. Access can read btrieve tables only if there are data definitions available, which are Btrieve .DDF files. If there are no ddf's, the processing archivist will need to create them from scratch with either a 3rd party tool such as BtSearch or use Actian's products. With both the BTR and DDF files and the Actian/Pervasive ODBC driver to set up ODBC DSNs the data can be read with MS Access.

Maintenance

- Updated quarterly
  - Changing risks such as format age
  - Revising available tools
  - Incorporating additional resources

- Ongoing research and conversations
  - Do our previous assumptions hold true?
  - Are we able to consistently apply risk analysis, within an acceptable margin of error, with our current methodology?

Using the Framework

- The NARA Digital Preservation Framework is publicly available: **https://github.com/usnationalarchives/digital-preservation**

- All community use and adaptive re-use is welcome, as is ongoing feedback.

Thanking our Team

- We want to acknowledge the cross-division team who contribute to this work:
  - Brett Abrams (Archivist, Electronic Records Division)
  - Criss Austin (Motion Picture Preservation Specialist, Special Media Division)
  - Dara Baker (Format Specialist, Office of Innovation)
  - Meg Guthorn (Chief, Still Pictures Division)
  - Michael Horsley (Electronic Record Format Specialist, Office of the Chief Records Officer)
  - Jana Leighton (Archivist, Electronic Records Division)
  - Andrea Riley (Supervisor, Operations Research and Support Team, Office of the Chief Records Officer)

Thank You

Leslie Johnston
leslie.johnston@nara.gov

Elizabeth England
elizabeth.england@nara.gov