

SOP for NARA Digital Preservation Framework

Table of Contents

Table of Contents	1
SOP Revision and Review History	2
SOP Purpose Statement and Scope	3
When does this SOP take effect?	3
Terms Used	4
NARA Acronyms and Terms	4
Non-NARA Acronyms and Terms	4
Infrastructure/Equipment	4
Computer Hardware, Software	4
Other Equipment and Supplies	5
Methodology	5
Preservation Action Plans: File Formats	5
File Format Identifiers Section	5
Links Section	5
Pause: Complete Risk Matrix	6
Proposed Preservation Actions Section	6
File Format Risk Matrix	7
Risk	7
Prioritization	9
Unspecified version	9
Preservation Action Plans: Categories	10
Document Title	10
Electronic Record/Surrogate Type Section	10
Significant Properties Section	10
NARA Transfer Guidance Section	11
NARA Public Access and Reference Section	11
Comments and Notes Section	11
Markdown Versions on Github	11
Change Log	12
Metadata	12
Control List	13
File Management	13
Quality Control and Updating the Change Log	13

SOP Revision and Review History

Version	Date Created/Updated	Who	Revision Description
0.1	10/22/2018	Leslie Johnston	Initial SOP Draft
1.0	2/15/2019	Criss Austin, Meg Guthorn, Leslie Johnston, Jana Leighton, Andrea Riley	Revised after review
1.1	6/25/2020	Elizabeth England	Revised in response to updating formatting and content for the Preservation Action Plans and File Format Matrix
1.2	4/26/2021	Elizabeth England	Revised threshold number of files in the holdings from 1,000 to 2,000 for inclusion in the Framework; expanded file management; updated terminology, template, SOP version numbering
1.3	10/7/2021	Elizabeth England	Revised SOP title; added information about dates for file formats in risk matrix; changed separator between multiple values from being a semicolon to a pipe
1.4	1/5/2022	Elizabeth England	Revised Proposed Preservation Actions Section to reflect updated controlled lists.
1.5	8/3/2023	Elizabeth England	Added new sections: Change Log, Control List, Ongoing Maintenance. Updated Authority section, WikiData URL information, and publishing instructions.
1.6	5/16/2024	Hannah Wang	Updated Change Log section. Added Quality Control and Updating the Change Log section.
1.7	9/26/2024	Hannah Wang	Added Appendix A and Appendix B. Updated Risk Matrix and Publishing to GitHub instructions to reflect the newest version of the spreadsheet. Moved Risk Matrix scoring

			guidance to Appendix A. Updated file paths to reflect new Digital Preservation Shared Drive.
1.8	12/4/2024	Hannah Wang, Amanda May, Mackenzie Beasley	Reorganized Methodology section to place Preservation Action Plans for File Formats before Risk Matrix. Added maintenance activity for updating Preservation Action Plans for Record Categories to reflect new Transfer Guidance. Added more guidance to Appendix A. Added instructions for updating Markdown for Preservation Action Plans for Record Categories.

SOP Purpose Statement and Scope

Documenting decisions that guide digital preservation operations is a vital activity. A Preservation Action Plan is a document used to assess risks and lay out the steps to mitigate the risks to best preserve the holdings in question. A Plan must identify risks, prioritization, proposed preservation actions, necessary resources, and procedures. NARA has developed a Digital Preservation Framework containing documents to serve this need, documenting the risk associated with file formats in its File Format Matrix, and recording the decisions made about the preservation of file formats in its File Format Preservation Action Plans. This SOP covers the creation, review, and updating of both the Matrix and the Preservation Action Plans.

When does this SOP take effect?

- This guidance takes effect as of **December 2, 2024**.
- The SOP consists of **24** pages.
- This SOP about digital preservation supersedes and replaces version 1.7.

Terms Used

NARA Acronyms and Terms

Acronym	Definition
R	Research Services. NARA's Office of Research Services is the custodian for federal archival records that are stored, managed, and made available at several locations across the country, which provide services for the public to discover, locate, and use the records.
L	Legislative Archives, Presidential Libraries and Museum Services
P	Presidential Libraries

Non-NARA Acronyms and Terms

Acronym	Definition
Preservation Action Plan	A Preservation Action Plan is a document used to assess risks and lay out the steps to mitigate the risks to best preserve the holdings in question. A Plan must identify risks, prioritization, proposed preservation actions, necessary resources, and the steps required. ¹

Infrastructure/Equipment

Computer Hardware, Software

Tool	Purpose	NARA IT Supported	Approved for NARAnet	Purchase Yr.
Google Sheets	File Format Matrix; Preservation Action Plans for File Formats; Control list for NARA IDs; Matrix Weights	Yes	Yes	
Google Docs	Preservation Action Plans for Categories	Yes	Yes	
Google Spaces	Collaboration and coordination	Yes	Yes	
GitHub	Dissemination to the public	Yes	Yes	

Other Equipment and Supplies

Tool	Purpose	NARA Supported	Purchase Yr.
N/A			

¹Digital Preservation Coalition. "Preservation Planning," n.d.
<https://www.dpconline.org/handbook/organisational-activities/preservation-planning>.

Methodology

Preservation Action Plans: File Formats

Data entry in the File Formats Preservation Action Plans must be precise in order for the later Linked Data generation to be done properly. Ensure you follow existing patterns (capitalization, using Pipes | as separators where applicable, etc.) in the Plans spreadsheet, and the instructions here.

The columns are as follows:

File Format Identifiers Section

- **File Format Name:** Provide the file format name, including applicable Publisher and version/date information, e.g., Microsoft Access 97, or Adobe InDesign Document CS. Follow a consistent format name pattern across related formats.
- **Extension(s):** List all applicable file extensions. A minimum of one is required although several file extensions may be associated if it is a compound object or container files such as a GIS object. The same file extension may be repeated on different rows for multiple versions across time. Separate multiple extensions with pipes | without spaces around the pipe.
Example: xls|xlsx
- **Category/Plan(s):** Enter the Record Type/Plan(s) for which the file format belongs. Separate multiple record types with pipes | without spaces around the pipe.
- **NARA Format ID:** Go to the Control List and add a new row at the bottom of the spreadsheet. The NARA Format ID for this format is the next available ID.
- **MIME type(s)** Enter the relevant MIME type(s) for the format. You can include unofficial ones if no registered MIME types exist. Separate multiple MIME types with pipes | without spaces around the pipe.

Links Section

It is not expected that each of these fields can be filled out; in many cases a relevant link doesn't currently exist. *Note: URLs, particularly for format specifications, sometimes include spaces due to pointing to a PDF or other file that has a space in the filename. Spaces are encoded in a URL as %20. This will cause issues later with the Linked Data and **URLs with spaces/%20 should be avoided**. Consult with the team in the Digital Preservation Framework Google Space for alternatives.*

- **Specification/Standard URL** *Include a link to the file format specification/standard. If one cannot be located, you can include a reverse-engineered specification, if that can be located. Note this should be a specification for the file format, not for related software. **Add the specification to the file format specifications library, using the process documented in the File Format Specifications Library SOP.***
- **[PRONOM URL](#)** *Include a link to the relevant PRONOM entry. Note this link should end in the PRONOM Persistent Unique Identifier (PUID), and may be different from what appears in your address bar, depending on how you navigated to the entry. If a PUID has been deprecated, include the link to the referenced PUID.*
- **[Library of Congress URL](#)** *Include a link to the relevant LC Sustainability of Digital Formats entry. If one cannot be found for the individual format, you can include a link to the format family entry if that exists. Links to related FDDs that are not part of the same family can be added to "Other URL."*

- **[British Library URL](#)** Include a link to the relevant British Library PDF, linked from the Digital Preservation Coalition's File Formats Assessments.
- **[WikiData URL](#)** Include a link to the relevant WikiData entry. If one cannot be found for the individual format, you can include a link to the format family entry if that exists. Do not include WikiData URLs for related software, encodings, etc. - only file formats.
- **[ArchiveTeam URL](#)** Include a link to the relevant ArchiveTeam entry. These are often for the format family, and not individual versions.
- **[ForensicsWiki URL](#)** Include a link to the relevant ForensicsWiki entry.
- **[Wikipedia URL](#)** Include a link to the relevant Wikipedia entry. In some cases, this may be for software associated with the file format.
- **[File Format Wiki \(wiki.fileformat.com\)](#)** This is optional, and may be more useful for some Record Types over others.
- **Other URL** This is optional, if there is another URL that fills a gap in understanding the format being described. This is useful for formats that do not have many other links. Possible other websites for research include:
 - **Fileextensions.org** (be cautious, this has lots of links to downloads/shareware, etc):
<https://www.file-extensions.org/>
<https://www.file-extensions.org/extensions/common-file-extension-list>
 - **dotwhat** (same caveat as fileextensions.org): <http://dotwhat.net/>
 - **FME formats matrix** (useful to see what formats FME supports):
<https://engage.safe.com/fme/formats-matrix/>
 - **Global Mapper supported formats**:
https://www.blumarblegeo.com/knowledgebase/global-mapper-22/File_Formats.htm

Pause: Complete Risk Matrix

At this point, you will need to calculate the Risk Level in the Risk Matrix before moving on to the next section in the Preservation Action Plans. See [File Format Risk Matrix](#).

Proposed Preservation Actions Section

- **Notes** This is optional, if there is additional information you want to provide that isn't captured elsewhere. For any unspecified version entry, include an explanatory note. See JPEG for example.
- **Risk** Include the overall Risk (Low Risk, Moderate Risk, or High Risk) from the Matrix.
- **NARA Preservation Action** Controlled List. The options are:
 - **Identify Version** Use when the most imminently needed preservation action is to identify the version(s) of a format, as version-based identification has not been performed and the longer-term preservation action may be version dependent (e.g., some versions of Microsoft Word are retained, and some versions are transformed. Therefore the preservation action for NF00659, Microsoft Word unspecified version, is Identify Version).
 - **Transform**
 - **Retain**
 - **Retain for Future Assessment** Use when further research is required, unknown if it will be retained in current format or transformed.
- **NARA Proposed Preservation Plan** Semi-controlled list. Use the following options, with additional description as necessary:
 - **Depends on Version** Use when the Preservation Action is Identify Version.
 - **Further research is required**
 - **Transform to a TBD format**

- **Transform to [insert format name]** *Use when target format is known, and preferably transformation tool(s) are already available to NARA.*
- **Retain** *Use when the Preservation Action is Retain.*
- **Retain but extract files from the container** *Use when the format is a container such as TAR or ZIP archive.*
- **Description and Justification** *You must supply a Justification for the Preservation Action(s) selected, which should be no more than 2 brief paragraphs. This can be as brief as a single sentence. Provide additional, brief description of the format as necessary.*
- **NARA Preferred Processing and Transformation Tool(s)** *Provide the name of the preferred tool(s) used within NARA to review the records in this format and prepare them for ingest. Provide the name of the preferred tool(s) used within NARA to validate and/or transform records in this format into new formats for preservation. If there is a known, preferred software package that should be used or recommended for working with records in this format, document the software here. Separate tools with semicolons.*
 - *If transformation tools are not currently available to NARA, include “Procure and/or develop tools”*
 - *If you are unsure what tools are available within NARA for working with a format, consult with the team in the Digital Preservation Framework Google Space.*

File Format Risk Matrix

Editable spreadsheet: Digital Preservation Framework / Preservation Action Plans / Editable Current Versions of Docs / EDITABLE_NARA_File_Format_Risk_Matrix_2024XXXX

The Risk Matrix has three sheets: Input, Numbered, and Labeled. As you are researching new formats or updating existing entries, you will only be interacting with the Input sheet. The Numbered and Labeled sheets are locked and auto-generated based on the values in the Input sheet. If any issues are encountered with the formulas in the Numbered and Labeled sheets, consult the team in the Digital Preservation Framework Google Space.

Risk

In the File Format Matrix you will answer questions about sustainability factors that impact the ability to preserve those formats. The sustainability factors fall into eight categories, each of which is weighted differently as it relates to the level of risk/sustainability and, to the extent that it can be identified to be taken into account, cost. Categories and questions have different weights related to their impact (positive or negative) on sustainability of a format and therefore its risk level. The weight assigned to each factor is documented in a weights spreadsheet, available in Drive.

Instructions:

1. Open the editable version of the Risk Matrix and go to the Input sheet. The Numbered and Labeled sheets are locked and auto-generated based on the values in the Input sheet.
2. Create a Row for the file format. Each File Format variation, such as a *unique* release/version, must have its own row in the sheet. *Note that generally, NARA considers a “unique” release/version to correspond to one PRONOM ID (PUIID). For example, Microsoft Word for MS-DOS versions 1-4 share [one PUIID](#) and correspond to one NARA Format ID.*
 - a. The general guideline is to create an entry for any file format variation for which the count is 2,000 or higher. There are many formats for which the variant/version information is currently unknown. In these cases, create an “unspecified version” entry

(see JPEG unspecified version, NF00641 for example). The count information is gathered from the Holdings Profile. It is not necessary to make an unspecified version entry for formats for which only a handful of files are identified for existing entries. See [Unspecified version](#) for instructions on creating these entries, as they differ from other entries in the Risk Matrix.

3. *Guidance Rating*: Put an X in the appropriate column if the format is on the NARA Preferred or Acceptable Transfer Guidance list.
4. *File Format Identifiers*: Previously described in the [Preservation Action Plans: File Formats](#) section. It is recommended to copy + paste the File Format Identifiers between the Matrix and File Formats Plans, rather than manually typing, to ensure consistency.
5. *Sustainability Factors*: For each of the following categories, research the format and enter the value that best addresses the sustainability of the format. Do not overthink this; the Matrix is not meant to be exhaustive and perfect. This is a living document that will be updated over time.
 - **Disclosure** *The degree to which complete specifications and tools for validating technical integrity exist and are accessible to those creating and sustaining digital content.*
 - **Adoption** *The degree to which the format is already used by the primary creators, disseminators, or users of information resources. This includes use as a primary or “master” format, for delivery to end users, and as a means of interchange between systems.*
 - **Transparency** *The degree to which the digital representation is open to direct analysis with basic tools, including human readability using a text-only editor.*
 - **Self-Documentation** *Self-documenting digital objects contain basic descriptive, technical, and other administrative metadata.*
 - **External Dependencies** *External dependencies refers to the degree to which a particular format depends on particular hardware, operating system, or software for rendering or use and the predicted complexity of dealing with those dependencies in future technical environments.*
 - **Licensing and Patents** *Degree to which the ability of archival institutions to sustain content in a format will be inhibited by licenses or patents.*
 - **Encryption and Rights Management** *Implementation of mechanisms such as encryption that negatively impact and prevent the preservation of content by a trusted repository.*
 - **Additional guidance**: See [Appendix A: Risk Matrix Scoring Guidance](#).

Prioritization

Next you will answer questions about NARA-specific factors that impact the prioritization for preservation actions.

Instructions:

1. *Prevalence*: Enter the count for this format currently identified in the NARA holdings across all NARA preservation systems.
 - If you cannot find the count, ask the Senior Digital Preservation Analyst to track it down for you.

- Note that the raw counts are removed from the Numbered and Labeled versions of the Risk Matrix before making them publicly available. See [Publishing to GitHub](#).
- Feasibility: The answer to this question often comes from the applied experiences of processing archivists at NARA, who are the best resources. Digital Preservation can assist in research on tools in the marketplace that have been identified or are in use at NARA.

Unspecified version

A number of files in NARA's holdings have only been identified by their extensions. File formats in the Risk Matrix that have “unspecified version” in their name currently serve as catch-alls for these files, until they can be identified more precisely.

In the Risk Matrix, entries for unspecified versions are largely left blank, except for the risk rating, risk level, and prioritization questions. Many of the questions are impossible to answer for a format if you do not know the specific version, because the answers may vary depending on the version. These unspecified versions are instead assigned the same overall risk rating as the lowest-scoring format with that extension, assuming the highest level of risk.

Instructions:

- In the Risk Matrix Input sheet, enter Guidance Rating and File Format Identifiers information about the unspecified version (columns A-F).
- Leave all questions blank from question 1.1 to question 8.3.
- Enter the number of identified files in the combined NARA holdings and the feasibility
- Open the Numbered sheet, scroll to the corresponding file format family/group of extensions and find the lowest score for this extension.
 - In this example, the lowest score for bmp is 14.00

Numeric Risk Rating	Risk Level	NARA Format ID	Format Name	File Extension(s)
14.00	Moderate Risk	NF00441	Windows Bitmap 1.0 / Windows Device-Dependent Bitmap	bmp
14.00	Moderate Risk	NF00442	Windows Bitmap 3.0	bmp
14.00	Moderate Risk	NF00443	Windows Bitmap 3.x NT	bmp
14.00	Moderate Risk	NF00624	Windows Bitmap 4.0	bmp
16.00	Moderate Risk	NF00625	Windows Bitmap 5.0	bmp

- In the Input sheet, enter this score in the final column (“If this format is an unspecified version, enter the lowest score for a format with this extension here. Leave empty if the format is not an unspecified version”). This will populate the Numbered and Labeled spreadsheets with this value and the corresponding risk rating.

Preservation Action Plans: Categories

Editable documents: Digital Preservation Framework / Preservation Action Plans / Editable Current Versions of Docs / Category Plans

The granularity for these Plans is at the Category level because significant properties of electronic records are shared by a category of records, such as word processing files or email, not to specific formats such as Microsoft Word 97 or EML.

Document Title

Identify the Category in the document Title, with the document date and the date for the version of the Template that was used. Example:

Preservation Action Plan: Web Records

Approved: 09122018
Using Template 201808

Electronic Record/Surrogate Type Section

Provide a brief description of the Category, no more than two brief paragraphs.

Significant Properties Section

Provide brief context--no more than 3-4 short paragraphs--for the Significant Properties, including an explanation of which Properties are the most significant for the Category (Record Type).

Identify the Significant Properties for each category. It is expected that some record type categories will have no significant properties in one or more categories.

- **Appearance** *Properties related to the visible appearance of this record type, which contributes to the re-creation of the record content and are required to convey meaning, as well as properties of the content of this record type which must be conveyed in any migration. Example: font type, appearance features, color and size, and bit depth.*
- **Structure** *Properties that are required to retain the structure of the information contained in this record type, such as information that describes the relationship between two or more types of content, as required to reconstruct its performance. It may be applied to the intrinsic or extrinsic relationships contained in the performance.
Ex: logical properties, duration, character count, external file relationships (such as email with attachment or threaded messages, etc.*
- **Behavior** *Properties related to user interaction with this record type, which may include the interaction of the user with the software, or interaction with other sources of information, such as an external resource that affects the content, context, structure, or appearance of the resources. Behavior is a difficult property to preserve, as it is often tied to the capabilities of a particular software application and may be difficult to translate.*
- **Context** *Properties required to provide context about this record type or its relationship to other records of the same or different types, such as the environment in which the Context was created or that affects its intended meaning. Ex: Creator name, date of creation, description of the intellectual work, computer environment in which the record was created.*

NARA Transfer Guidance Section

Document the current Preferred and Acceptable file format(s) for this Category for transfer from agencies to NARA.

NARA Public Access and Reference Section

Formats for Public Access are those made available online through the National Archives Catalog.

Formats for Reference are defined as those made available to researchers upon direct requests for digital copies.

Comments and Notes Section

Enter any appropriate notes and comments about sources for significant properties and findings.

Markdown Versions on Github

https://github.com/usnationalarchives/digital-preservation/tree/master/Digital_Preservation_Record_Categories

First document any changes made to the Action Plans each quarter. To do so, select an action plan, and click the pencil on the top right to edit. Always save the edits and if it's a quarterly update, write in the extended description the date like "20241205 Update." No need to write an extended description if it's a typo or a small edit.

Tips for Markdowns:

 is the equivalent of hitting the "enter" button and creates a break or space
WORDING create bullet points, specifically used in tables.

Then make sure to upload the most current PDF copy into the PDF folder

(https://github.com/usnationalarchives/digital-preservation/tree/master/Digital_Preservation_Record_Categories/PreservationActionPlan_PDFs)

Change Log

Editable Google Doc: Digital Preservation Framework / Preservation Action Plans / Editable Current Versions of Docs / EDITABLE_NARA_Framework_Change_Log_2023XXXX

Keep track of the new formats added each quarter, and edits made to existing data, for publishing in the [Change Log](#) in GitHub. The Change Log does not need to include every link that was updated, or available tool, but should cover the following types of changes:

- Added
- Changed
- Deprecated
- Fixed

Categorize these changes across the following categories:

- Record Categories
- General changes to both the 1) Risk and Prioritization Matrix and 2) File Formats Preservation Action Plans
- Risk and Prioritization Matrix

- File Formats Preservation Action Plans

Make sure to capture:

- Formats added (NARA identifier, format name)
- Formats deprecated (NARA identifier, format name, reason for deprecation)
- Format names changed
- Changed risk levels (such as from Low to Moderate Risk, do not need to track small changes in numeric scores)
- Changed preservation plans (such as from Transform to Retain)

Publish the changes in [markdown](#) in the Change Log on GitHub. Make sure to increment the version for each release and add the date to the Change Log. Versions should be written in a format that complies with [Semantic Versioning](#) guidelines: MAJOR.MINOR.PATCH.

For example, [2.13.0] - 2023-12-20 is the 13th minor version of the second major release, published on December 20, 2023.

Metadata

All metadata is embedded in the structure of the Matrix and Plans.

Plans must include the date that the current version of the Plan is complete and supersedes any previous versions.

Control List

The Control List is an internal-only spreadsheet, located in the Digital Preservation Unit Shared Drive: *Digital Preservation Framework / Preservation Action Plans / Editable Current Versions of Docs*. The purpose of the Control List is to track the assigning of NARA IDs, and it includes current and deprecated IDs.

File Management

A copy of the most current version of each file should be stored in the Digital Preservation Unit Shared Drive: *Digital Preservation Framework / Preservation Action Plans / Editable Current Versions of Docs*. Completed releases should include the date in the filenames. We want to keep versioned copies to trace the history of our preservation decisions and justifications.

- NARA_File_Format_Risk_Matrix_Weights_YYYYMMDD (spreadsheet)
- NARA_Deprecated_IDS_YYYYMMDD (spreadsheet)
- control_list_NARA_IDS (spreadsheet used for assigning new IDs, not published to GitHub)
- NARA_PreservationActionPlan_FileFormats_YYYYMMDD (spreadsheet)
- INTERNAL_NARA_File_Format_Risk_Matrix_YYYYMMDD (spreadsheet)
- NARA_File_Format_Risk_Matrix_YYYYMMDD (spreadsheet)
- NARA_PreservationActionPlan_WebRecords_YYYYMMDD (text document, follow this naming convention for all category plans)

Preservation Actions Plans must be reviewed and updated as needed on a quarterly basis at a minimum. Trigger events for the creation of a new plan include, but are not limited to, the addition of new Categories to the NARA Transfer Guidance, or updates to the Holdings Profile. When updating the Plans or Risk Matrix, create a **copy** of the newest completed release.

Quality Control and Updating the Change Log

Before publishing a new release, it can be helpful to compare the current versions of both the Risk Matrix and the Preservation Action Plans using a diffchecker (such as <https://www.diffchecker.com/>):

1. Save the current Risk Matrix or Preservation Action Plan locally as a CSV file. Make sure that the spreadsheet is sorted alphabetically by Format Name.
2. Download the last completed release of the Risk Matrix or Preservation Action Plan from *Preservation Action Plans / Completed Releases*. If you are comparing the Risk Matrixes, download the INTERNAL version of the Risk Matrix. Save this file as a CSV.
3. Upload both versions to a diffchecker (using the last completed release as the original and the current spreadsheet as the updated version) and do a visual scan of the changes. Look for potential copy/paste errors and typos. You can also use the diffchecker to determine significant changes that need to be added to the Change Log.
 - a. You can also use the comparing changed risk spreadsheet to compare the risk levels of the old and new releases

It is also important to compare the fields that should be identical between the Risk Matrix and the Preservation Action Plans (NARA Format ID, Format Name, File Extension(s), Category/Plan, Risk Level):

1. Open the comparing identical fields spreadsheet and follow the instructions documented in the first sheet.
2. If you find mismatches (which will be indicated as FALSE values in the Comparison sheet), update the Risk Matrix and/or the Preservation Action Plans so that these values match. If you are unsure how to amend the issue, consult the team in the Digital Preservation Framework Google Space.
3. Once you have made these updates, repeat the process to make sure that there are no remaining mismatches.

Publishing to GitHub

When ready for release, create a new folder in *Preservation Action Plans / Completed Releases* and name the folder with the month and year of the release, such as 2021-03. Any files which should not be uploaded to GitHub, such as working files and the internal-only version of the Matrix with raw counts present, should be clearly labeled as INTERNAL.

To create the external version of the Matrix:

1. Download the Matrix spreadsheet as an .xlsx file
2. In the “Labeled” sheet:
 - a. Copy + paste all values as **values only**
 - b. Scroll to the bottom of the entries and delete all empty rows
 - c. Delete column AT (Number of identified files)
3. Repeat step 2 for the “Numbered” sheet
4. Delete the “Input” sheet
5. Save the file as NARA_File_Format_Risk_Matrix_YYYYMMDD.xlsx

6. Save each sheet as NARA_File_Format_Risk_Matrix_YYYYMMDD_Numbered.csv and NARA_File_Format_Risk_Matrix_YYYYMMDD_Labeled.csv

The Matrix and File Formats Preservation Action Plans spreadsheets should be released as XLSX and CSV files. The Category Preservation Action Plans should be released as Markdown and PDFs. The Change Log should be released as Markdown.

Include an external-facing version of this SOP for GitHub. Links and file paths to NARA-internal systems (ICN, Google Drive, NARA at Work) and the appendices should be removed for the GitHub copy, and it should be released as a PDF.

Work with the team to coordinate publication on GitHub, and timing with publishing corresponding Linked Data. When uploading a new version of a file to GitHub, the new version can replace the old one. The only exception to this is the Change Log, a continuous Markdown file that is updated each quarter.

Ongoing Maintenance

In addition to adding new entries to the Framework over time, there is ongoing maintenance to be done.

Preservation Action Plans: File Formats

- Checking/replacing broken links (this may require searching for the original link in the Internet Archive's Wayback Machine)
- Filling in gaps when new links/information is available (PRONOM, Wikidata, LOC, etc.)
- Checking for currency of information about NARA practices and tools

Risk Matrix

- Reviewing data entry for consistency (such as checking across format variants of one format family)
- Ensuring Feasibility scores accurately reflect NARA's current practices and tools
- Checking for recent updates to format specifications
- Reflecting any updates to Preferred or Acceptable formats in the Transfer Guidance
- Updating the number of identified files in NARA's holdings *Note: this is visible internal only*
- Annually, at the end of each calendar year (December quarterly release), the formulas used to calculate format age need to be updated (columns BC and BE). The year used should be the coming calendar year. *Note: this may cause some formats to increase from Low to Moderate Risk, or Moderate to High. When done with all updates, copy + paste the Risk Level values to the Preservation Action Plans spreadsheet.*

Preservation Action Plans: Record Categories

- Reflecting any updates to Preferred or Acceptable formats in the Transfer Guidance

SOP Update Strategy

This SOP will be updated every odd numbered fiscal year during the first quarter **or** if there is a trigger event such as changes to business processes, to staffing, and/or to systems.

When the SOP is updated, include the most up-to-date version in the next round of publishing to GitHub.