

SOP for NARA Digital Preservation Framework

Table of Contents

| | |
|--|----------|
| Table of Contents | 1 |
| SOP Revision History | 3 |
| SOP Purpose Statement and Scope | 3 |
| Authority for Creating the SOP | 3 |
| When does this SOP take effect? | 4 |
| Terms Used | 4 |
| NARA Acronyms and Terms | 4 |
| Non-NARA Acronyms and Terms | 4 |
| Infrastructure/Equipment | 5 |
| Computer Hardware, Software | 5 |
| Other Equipment and Supplies | 5 |
| Methodology | 6 |
| File Format Matrix | 6 |
| Sheet One: Risk Matrix | 6 |
| Sheet Two: Prioritization Matrix | 10 |
| Preservation Action Plan | 11 |
| Document Title | 11 |
| Electronic Record/Surrogate Type Section | 12 |

| | |
|---|-----------|
| Essential Characteristics Section | 12 |
| NARA Transfer Guidance Section | 13 |
| Comments and Notes section | 13 |
| Associated File Formats section | 13 |
| File Format name | 13 |
| Documentation section | 13 |
| Risk and Prioritization Analysis Section | 13 |
| Proposed Preservation Plan Section | 13 |
| Preferred Processing and Transformation Tools Section | 14 |
| Preferred Viewer/Access Software section | 14 |
| Metadata | 14 |
| File Management | 14 |
| SOP Update Strategy | 15 |
| Appendix A: Details for Matrix Sheet Columns | 16 |

SOP Revision History

| Version | Date Created/Updated | Who | Revision Description |
|---------|----------------------|--|----------------------------|
| 0.1 | 10/22/2018 | Leslie Johnston | Initial SOP Draft |
| 0.2 | 2/15/2019 | Criss Austin, Meg Guthorn Leslie Johnston, Jana Leighton, Andrea Riley | Revised after review |
| 0.3 | 8/27/2019 | Leslie Johnston | Revised for Public Release |

SOP Purpose Statement and Scope

Having documented decisions that guide digital preservation operations is a vital activity. NARA has developed a Digital Preservation Framework 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. 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. This SOP covers the creation, review, and updating of both the Matrix and the Preservation Action Plans.

Authority for Creating the SOP

- NARA Notice 2017-111: Digital Preservation Strategy and Digital Preservation Group: <https://work.nara.gov/notices/2017/111>
- NARA Digital Preservation Strategy (2017) identifies the need for “Documentation of Standards and Procedures,” which includes SOPs as internal documentation of and guidance for operations. https://work.nara.gov/files/notices/DPG_Charter_and_DP_Strategy.pdf
- NARA Notice 2018-066: NARA’s 2018–2022 Strategic Plan and FY 2019 Budget Request: <https://www.nara-at-work.gov/notices/2018/066>

When does this SOP take effect?

- This guidance takes effect as of **January 1, 2019**.
- The SOP consists of **12** pages.

Terms Used

NARA Acronyms and Terms

| Acronym | Definition |
|---------|------------|
| N/A | |

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. |

Infrastructure/Equipment

Computer Hardware, Software

| Tool | Purpose | NARA IT Supported | Approved for NARAnet | Purchase Yr. |
|---------------|--|--------------------------|-----------------------------|---------------------|
| Google Sheets | File Format Matrix; Preservation Action Plan Overview; Essential Characteristics Matrix | Yes | Yes | |
| Google Docs | Preservation Action Plans | Yes | Yes | |

Other Equipment and Supplies

| Tool | Purpose | NARA Supported | Purchase Yr. |
|-------------|----------------|-----------------------|---------------------|
| N/A | | | |

Methodology

File Format Matrix

Sheet One: Risk Matrix

In The File Format Risk Matrix you will answer questions about sustainability factors that impact the ability to preserve those formats. The sustainability factors fall into nine 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.

- Highest Impact Categories
 - Positive: Disclosure, Adoption
 - Negative: Hardware Dependency, Software Dependency, Age of Format
- Highest Impact Questions
 - Positive: +2 for several questions in Disclosure, Adoption, Self-documentation, Software Dependency, and Technical Protection Mechanism categories.
 - Negative: -4 for Format Age and Hardware Dependency questions.

Instructions:

1. 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.
2. *Guidance Rating*: Put an X in the appropriate column if the format is on the NARA Preferred or Acceptable Transfer Guidance list
3. *Overall Risk Rating*: This will be auto-completed at the end of data entry.
4. File Format Identifiers:
 - Provide the file format name.

- 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.
 - Pick from one of the Categories in the Drop Down List. New Categories can be added upon request as appropriate.
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 or over-research 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 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.*
 - **Format Age** *Risk Factors related to legacy and future NARA holdings identified by the NARA Digital Preservation Group.*
- **Additional guidance:** When answering Risk Matrix questions, the following guidance may be of use. If a topic is not mentioned here, use your best judgement when filling out the risk matrix. It is acceptable and unavoidable to have factors for which the value is “0” because there is no applicable answer or the answer cannot be determined. This is especially true for proprietary formats for which there is no open specification or other resource to confirm information, specially the status of patent/licensing issues. The Licensing/Patents and Encryption/Rights Management sections are the most likely to have “0” values.

- *What is an “published open specification?”* An open specification is a complete technical format specification created and controlled, in an open and fair process, by an association or a standardization body intending to achieve interoperability and interchangeability; rights, licenses, and patents associated with the specification must be clearly communicated.
- *How do you know if there are available tools that can validate the technical integrity of a file encoded in this format against the published specification?* Check the PRONOM database, JHOVE, and JHOVE2 to determine if the format is listed and includes a format signature, which is necessary for validation.
- *How do you know whether the specification been approved and published by an internationally recognized standards body?* If the standard comes from NISO, ISO, ANSI, WC3, IANA, another national or international standards organization, or an organization dedicated to a specific format, such as the PDF Association, it meets the criteria.
- *How do you determine whether the specification is complete and accurate?* Complete specifications should include information about file extensions, header structure, the magic number, data structure including byte patterns, extended attributes, and required or optional external linkages. File format Identifiers (FFIDs) are relatively new and may not be assigned/included. Accuracy is difficult to judge, but if you find other format-related resources citing the specification, it is likely to be complete and authoritative.
- *What constitutes common use in the federal government?* The answer is yes if NARA has added more than 5,000 files of this type to its holdings. This is a somewhat arbitrary threshold, the same that we have used to identify statistically significant formats for risk assessment.
- *What constitutes “multiple renderers?”* There should be 3 or more commercial, freeware, or open source tools that can render files of this type.
- *How do you know if the archives or library communities identify the format as one they prefer for creation and transfer of permanent materials?* Confirm that the format is on the NARA Transfer Guidance, the Library of Congress Recommended Formats Statement, or guidance from Libraries and Archives Canada, The National Archives UK, or the National Archives of Australia.
- *How can you determine whether the format relies on standard character or other encoding methods such as IEEE notation?* Hopefully any technical specification for a format or preservation description of the format will include this information. This may not always be discoverable.
- *How do you determine whether the software used to create the format is supported by current computing environments?* Documentation of software generally includes information about the environment(s) in which it runs, e.g. Win95, MacOS7, Windows 7/10, etc. An environment may be considered current, e.g. still commercially supported by the original company or by a 3rd party, but a separate issue is whether it is available at NARA for use in processing. Determining the availability of an environment will often require consultation with NARA IT on an individual facility and unit basis to determine whether we have matching computing capabilities and support, and/or whether we can obtain appropriate software/hardware and support.

- *What is considered “descriptive metadata?”* Metadata that describes a record for purposes such as discovery and identification, including agency name, record series, names, places and dates. Many formats allow some level of descriptive metadata to be embedded in the file headers.
- *What is considered “technical metadata?”* Technical metadata documents the technical environment in which a file was created and used, including hardware, software, operating systems, and specifications for the file (pixel height/width, encoding codec, etc). This is generally automatically written into the file header when it is created and when it is updated.
- *What is considered “administrative metadata?”* Administrative metadata has some commonalities with descriptive metadata--such as authority for the transfer/accessioning and the provenance of the record--that is used to manage the files, but also includes metadata related to rights and restrictions.
- *How robust should the metadata be for an accurate file analysis?* This varies by record type. The definition of robustness is the ability to determine if the file is the format it purports to be, that it is well-formed, and can be rendered or transformed into a renderable format, and that there is sufficient descriptive metadata to make the records accessible.
- *Does the format require a specific hardware environment, such as a specific graphics card, chipset, or memory requirements, to process or interact with it?* This applies most often to special media and sometimes digital still images, as some formats are specific to the carrier media they are stored on.
- *Does the format rely on specific computing operating system(s) to render or view files?* Some formats are limited to the MacOS, Windows, DOS, or Linux/Unix environments because there is no software that supports the format available for other environments.
- *How do you know about patent claims, their status, and whether there are any associated fees?* Unless a format has its own web site or page documenting its history, the best resources to research this are the Library of Congress Format Sustainability site, Wikipedia, Wikidata, and the ArchiveTeam file format site. Answers to this question may not be easily discovered.
- *What is considered “robust encryption?”* “Robust” encryption involves mathematically calculated cipher/key pairs that are used to lock and unlock files, as opposed to simple password locking of files. Encryption is most commonly used for sensitive or classified records, to protect PII (such as user accounts or personnel data), or intellectual property (software code). Some formats natively support encryption of their file contents. Some encrypted formats are container formats that can hold other types of files.
- The equivalent for “human readable” for non-text files is that we must be able to run commonly available tools (ubiquitous?) that identify the format and recognize it as a format that can be rendered/played.

- **Resources for researching formats:**

- NARA Table of Preferred and Acceptable File Formats:
<https://www.archives.gov/records-mgmt/policy/transfer-guidance-tables.html>
- Library of Congress (LOC) Sustainability of Formats site: <https://www.loc.gov/preservation/digital/formats/>
- British National Archives PRONOM database:
<http://www.nationalarchives.gov.uk/PRONOM/Format/proFormatSearch.aspx?status=new>
- ArchiveTeam File Format site: http://fileformats.archiveteam.org/wiki/File_Formats
- Wikidata File Format project:
https://www.wikidata.org/wiki/Wikidata:WikiProject_Informatics/File_formats/Lists/File_formats
- ForensicsWikie file formats: https://forensicswiki.org/wiki/Category:File_Formats
- 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/>
- Wikipedia: <http://wikipedia.org/>

5. Summary: This will be auto-completed after all numeric values are assigned.

Sheet Two: Prioritization Matrix

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

Instructions:

1. Each File Format on the Risk sheet must have a corresponding row on the Prioritization sheet.
2. File Format Identifiers: Copy the Identifiers from the Risk sheet when you add the row.
 - Short name and an official name.
 - All applicable file extensions.
 - Category.

3. Need: Copy the Risk Level and calculated Risk numeric value from the Risk sheet.
4. Prevalence: Enter the current count for this format currently in the NARA holdings across all NARA preservation systems.
The percentage in the holdings will be automatically calculated from the count.
5. Feasibility: The answer to this question comes from the applied experiences of processing archivists in R and L, 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.
 - No acceptable tools available in the marketplace=-5
 - Acceptable tools exist but NARA does not have them=-3
 - Migration already performed at NARA=3
 - Preferred/Acceptable Format as per Guidance or practice where no migration is needed=5
6. Transfer Guidance Status: Copy the Transfer Guidance (Preferred or Acceptable) from the Risk sheet.
7. Total: The numeric rating will be automatically calculated.

Preservation Action Plan

The granularity for Plans is at the Record Type level, not the file format level, because essential characteristics of electronic records are shared by a record type, such as word processing files or email, not to specific formats such as Microsoft Word 97 or EML.

Document Title

Identify the Record Type 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

Template 2018008

Electronic Record/Surrogate Type Section

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

Essential Characteristics Section

Provide brief context--no more than 3-4 short paragraphs--for the Essential Characteristics, including an explanation of which Characteristics categories are the most significant for the Record Type.

Identify the Essential Characteristics for each category. It is expected that some Record Types will have no essential characteristics in one or more categories.

- **Appearance** *Characteristics 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 characteristics 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** *Characteristics 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** *Characteristics 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 characteristic to preserve, as it is often tied to the capabilities of a particular software applications and may be difficult to translate.*
- **Context** *Characteristics 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 Record Type, and the Delivery format for the National Archives Catalog, plus the preferred format used to provide reference services.

Comments and Notes section

Enter any appropriate notes and comments about sources for essential characteristics and finding.

Associated File Formats section

Create a full analysis section for each individual format. Separate each with a page break to distinguish between the sections.

File Format name

The Name of the specific individual format being analyzed.

Documentation section

Record the online resources or other documentation that was used to document the format analysis, in the format: Information source or author(s), page or document title (date of publication if known) .

Risk and Prioritization Analysis Section

Enter the Risk level, numeric risk rating, and numeric prioritization rating from the Risk and Prioritization Matrix.

Proposed Preservation Plan Section

Select the current Recommended Action(s) from the available options. You must supply a Justification for the option(s) selected, which should be no more than 2 brief paragraphs. This can be as brief as a single sentence.

Preferred Processing and Transformation Tools Section

Provide the name of the preferred tool(s) used to review the records in this format and prepare them for ingest.

Provide the name of the preferred tool(s) used to transform records in this format into new formats for preservation and, if a different format is required, for public access and reference use.

Preferred Viewer/Access Software section

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.

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.

File Management

Preservation Actions Plans must be reviewed and updated as needed on a quarterly basis at a minimum. Trigger events for the creation of new plan include the addition of new Record Types to the NARA Transfer Guidance.

New Plans must be new, dated versions, not edits to the existing Plans. We want to keep versioned copies to trace the history of our preservation decisions and justifications. Preservation Action Plan file naming must follow the convention:

PreservationActionPlan_RecordType_YYYYMMDD

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, staffing, and/or systems.

Appendix A: Details for Matrix Sheet Columns

| Disclosure | | | | | |
|--|--|---|--|---|---|
| Disclosure refers to the degree to which complete specifications and tools for validating technical integrity exist and are accessible to those creating and sustaining digital content. | | | | | |
| 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 |
| -1 = Yes, 0 = N/A or Unknown, 2 = No | 2 = Yes, 0 = N/A or Unknown, -1 = No | 2 = Yes, 0 = N/A or Unknown, -1 = No | 2 = Yes, 0 = N/A or Unknown, -2 = No | 2 = Yes, 0 = N/A or Unknown, -1 = No | Sum of Columns I-M |

| Adoption | | | | | |
|--|--|---|--------------------------------------|---|---|
| Adoption refers to the degree to which the format is already used by the primary creators, disseminators, or users of information resources. This includes use as a master format, for delivery to end users, and as a means of interchange between systems. | | | | | |
| Is the file format commonly used to create or maintain permanent records within the federal government? | Is the file format commonly used outside the federal government? | Is the format actively maintained and updated by an organization, individual, or community? | Are multiple renderers available? | Have the archives or library communities identified the format as one they prefer for creation and transfer of permanent materials? | Total Adoption Score: Highest possible score = 10; Lowest possible score = -6 |
| 2 = Yes, 0 = N/A or Unknown, -2 = No | 2 = Yes, 0 = N/A or Unknown, -1 = No | 2 = Yes, 0 = N/A or Unknown, -1 = No | 2 = Yes, 0 = N/A or Unknown, -1 = No | 2 = Yes, 0 = N/A or Unknown, -1 = No | Sum of Columns O-S |

| Transparency | | | | | | | |
|--|---|--|---|--|---|--|--|
| Transparency refers to the degree to which the digital representation is open to direct analysis with basic tools, including human readability using a text-only editor. | | | | | | | |
| Is the format human readable and can it be opened with a text editor? | Does the available specification provide enough detail to allow basic analysis of the file format with hex editors or other inspection tools? | Does the format rely on standard character or other encoding methods such as IEEE notations? | Is the source code of the software used to create the format available for little or no cost? | Is the software used to create the format supported by current computing environments? | Does the format require the use of compression? | Does the format support user-definable compression levels or other quality settings that affect essential characteristics? | Total Transparency Score. Highest possible score = 7; Lowest possible score = -7 |
| 1 = Yes, 0 = N/A or Unknown, -1 = No | 1 = Yes, 0 = N/A or Unknown, -1 = No | 1 = Yes, 0 = N/A or Unknown, -1 = No | 1 = Yes, 0 = N/A or Unknown, -1 = No | 1 = Yes, 0 = N/A or Unknown, -1 = No | -1 = Yes, 0 = N/A or Unknown, 1 = No | -1 = Yes, 0 = N/A or Unknown, 1 = No | Sum of Columns U-AA |

| Self-Documentation | | | | | |
|---|--|---|---|---|---|
| Self-documenting digital objects contain basic descriptive, technical, and other administrative metadata. | | | | | |
| Does the format support self-contained (embedded) descriptive metadata? | Does the format support self-contained technical metadata? | Does the format support self-contained administrative metadata? | Does the format metadata adhere to an international standard? | Is the format metadata robust enough for an accurate file analysis? | Total Self-Documentation Score. Highest possible score = 7; Lowest possible score = -5 |
| 1 = Yes, 0 = N/A or Unknown, -1 = No | 1 = Yes, 0 = N/A or Unknown, -1 = No | 1 = Yes, 0 = N/A or Unknown, -1 = No | 2 = Yes, 0 = N/A or Unknown, -1 = No | 2 = Yes, 0 = N/A or Unknown, -1 = No | Sum of Columns AC-AG |

| External Hardware Dependencies | | |
|--|---|--|
| The degree to which a format depends on particular hardware for processing and the predicted complexity of dealing with those dependencies in future technical environments. | | |
| Does the format require a specific hardware environment, such as a specific graphics card, chipset, or memory requirements, to process or interact with it? | Does the format require specific playback hardware (e.g., Blu-Ray, Audio CD, etc) to transfer the format to the NARA environment? | Total External Hardware Dependencies Score. Highest possible score = 4; Lowest possible score = -8 |
| -4 = Yes, 0 = N/A or Unknown, 2 = No | -4 = Yes, 0 = N/A or Unknown, 2 = No | Sum of Columns AI-AJ |

| External Software Dependencies | | | | |
|--|--|---|--|--|
| External dependencies refers to the degree to which a format depends on a particular operating system or software for processing or rendering or use and the predicted complexity of dealing with those dependencies in future technical environments. | | | | |
| Does the format rely on proprietary software to render or view files? | Does the format rely on plug-ins, scripts, etc. to render or view files? | Does the format rely on specific computing operating system(s) to render or view files? | Does the format rely on special tools to render or view files? | Total External Software Dependencies Score. Highest possible score = 4; Lowest possible score = -8 |
| -2 = Yes, 0 = N/A or Unknown, 1 = No | -2 = Yes, 0 = N/A or Unknown, 1 = No | -2 = Yes, 0 = N/A or Unknown, 1 = No | -2 = Yes, 0 = N/A or Unknown, 1 = No | Sum of Columns AL-AO |

| Impact of Patents and Licenses | | | | |
|---|--------------------------------------|---|---|--|
| Degree to which the ability of archival institutions to sustain content in a format will be inhibited by licenses or patents. | | | | |
| Is the format subject to patent claims that may impede the development of open source tools for opening and managing the files? | Have the patent claims expired? | Are there fees associated with the format as a result of patent claims? | Does the format have open source license terms? | Total Impact of Patents Score. Highest possible score = 4; Lowest possible score = -4 |
| -1 = Yes, 0 = N/A or Unknown, 1 = No | -1 = No, 0 = Unknown, 1 = Yes or N/A | -1 = Yes, 0 = N/A or Unknown, 1 = No | 1 = Yes, 0 = N/A or Unknown, -1 = No | Sum of Columns AQ-AT |

| Technical Protection Mechanisms | | | | | |
|---|--|--|--|--|--|
| Implementation of mechanisms such as encryption that negatively impact and prevent the preservation of content by a trusted repository. | | | | | |
| Does the format have capability to encrypt all or part of the resulting file? | Does the format require the use of encryption? | Does the format support robust encryption? | Can technical protection measures (e.g. digital rights management) be applied? | Does the format allow for embedded information such as watermarking? | Total Technical Protection Mechanisms Score. Highest possible score = 10; Lowest possible score = -6 |
| -1= Yes, 0 = N/A or Unknown, 2 = No | -2=Yes, 0=N/A or Unknown, 2 = No | -1= Yes, 0 = N/A or Unknown, 2 = No | -1= Yes, 0 = N/A or Unknown, 2 = No | -1= Yes, 0 = N/A or Unknown, 2 = No | Sum of Columns AV-AZ |

| Format Age | | | | |
|--|--|---|--|---|
| Age of the Format and Currency of the Format Specification as Risk Factors | | | | |
| When was the Format Specification First Created? | Risk Factor: Highest Possible score = 0; Lowest possible score = -4 | When was the Format Specification last updated? | Risk Factor: Highest Possible score = 0; Lowest possible score = -4 | Total Format Age Score. Highest possible score = 0; Lowest possible score = -8 |
| Year | 0=5 years old or less, -2 for 6-15 years, -4 for 16+ years | Year | 0=5 years old or less, -2 for 6-15 years, -4 for 16+ years | Sum of Columns BB-BE |

| Need: Need for Preservation Actions as measured by the Risk Matrix Value | | | Prevalence: Format Adoption Level as measured by Proportion of File Format in the Overall NARA Holdings | Feasibility: Ability to Convert (tools exist for conversion that does not alter content in unacceptable ways; NARA can perform acceptable transformations) Highest possible score=5, Lowest possible score =-5 | Transfer Guidance Status | | TOTAL |
|--|--|--|--|--|--------------------------|------------|---|
| Number calculated in the Risk Matrix | Number of identified files in the combined NARA holdings | Percentage of 1.5 billion total files in the NARA holdings | -5 for 0-2%, -6 for 3-4%, -7 for 5-6%, -8 for 7-8%, -9 for 9-10%, -10 for 11-12%, -11 for 13-14%, -12 for 15-16%, -13 for 17-18%, -14 for 18-20%; -15 for over 20% | No acceptable tools available in the marketplace=-5; Acceptable tools exist but NARA does not have them=-3; Migration already performed at NARA=3; Preferred/Acceptable Format as per Guidance or practice where no migration is needed=5; | Preferred | Acceptable | The lower the number, the higher priority for preservation actions based on risk level, percentage in the holdings, and the requirement/ability to perform format migrations. |