NARA Risk and Prioritization Matrix Proposed Updates

NARA uses the [Risk and Prioritization Matrix](#) to measure the preservation risk of digital file formats in our holdings and to assess formats we anticipate receiving in the future. By answering questions related to the ability to preserve and sustain a file format, we identify relative risk levels.

In an effort to make the Risk Matrix more standardized, scalable, and better aligned with our current practices, the NARA Digital Preservation Unit has performed a comprehensive review and update of all the questions and categories of questions. We are proposing several different kinds of changes, including removing and condensing similar questions, updating the relative weights of questions, enforcing logic pathways for certain types of questions, and adding new questions to address additional risk factors that have emerged over the past decade since the initial version of the Matrix was drafted.

For each category of questions, there is a table summarizing the questions, values assigned to each response, guidance on how the scorer should assign values, and a summary of updates made to the questions. Logic pathways are noted in italicized, bold question text. We have also documented questions that we plan to remove.

You can see examples of formats that have been rescored using these new questions, as well as the original scores, in the accompanying "NARA_File_Format_Risk_Matrix_Proposed_Updates_2024.xlsx" Excel file. You will also find a human readable "labeled answer" version of the updated Risk Matrix in this Excel file; this is not currently a version of the Risk Matrix that NARA publishes – please let us know if this is helpful to see and we can explore publishing it in the future.

We welcome your feedback on this document. Please submit feedback to digital-preservation@nara.gov or as a [GitHub issue](#) by May 20, 2024.

Disclosure

	Question	Values	Scoring Guidance	Updates
1.1	Is the format proprietary?	-1 = Yes or Unknown, 2 = No		Removed N/A option, scoring Unknown as a negative value
1.2	Does the format have a published open specification?	2 = Yes, -2 = No or Unknown		Removed N/A option, scoring Unknown as a negative value, giving heavier weight to No or Unknown
1.3	*If there is a published specification*, has it been approved and published by an internationally recognized standards body?	2 = Yes, 0 = N/A, -1 = No or Unknown	If answer to previous question is No or Unknown, N/A	Added conditional IF statement, scoring Unknown as a negative value, giving lighter weight to No or Unknown
1.4	*If there is a published specification*, has it been successfully used by a third party to open and/or create valid files?	2 = Yes, 0 = N/A, -1 = No or Unknown	For software and code, use N/A If answer to open specification question is No or Unknown, N/A	NEW QUESTION
	Total Score:	Highest possible: 8 (previously 10) Lowest possible: -3 (previously -6)		

Questions removed from Disclosure:
- Are there available tools that can validate the technical integrity of a file encoded in this format against the published specification? (similar question added to Transparency)
- Is the available specification complete and accurate? (too vague)

Adoption

	Question	Values	Scoring Guidance	Updates
2.1	Has the file format commonly been used to create or maintain permanent records within the federal government at any point in time?	2 = Yes, 0 = Unknown	If NARA has added more than 2,000 files to its holdings, the answer is yes	Updated tense in question wording (i.e., "is" to "has been"), removed N/A and No options
2.2	Is the format, or format family, actively maintained and updated by an organization or community?	2 = Yes, -1 = No, Unknown	If there's active maintenance and updates to the format family in the past 5 years, then give a positive score to older versions of the format family	Added "or format family", removed N/A option, scoring Unknown as a negative value
2.3	Have the archives or library communities identified the format as one they prefer for creation and transfer of permanent materials?	2 = Yes, -1 = No or Unknown	Consult the OPF International Comparison spreadsheet and other resources	Removed N/A option, scoring Unknown as a negative value
	Total Score:	Highest possible: 6 (previously 10) Lowest possible: -2 (previously -6)		

Questions removed from Adoption:
- Is the file format commonly used outside the federal government? (there is not enough data that can be consistently used to answer this question)
- Are multiple renderers available? (similar question 6.1 added to External Software Dependencies)

Transparency

	Question	Values	Scoring Guidance	Updates
3.1	Is the format human readable and can it be opened with a text editor?	1 = Yes, -1 = No or Unknown		Removed N/A option, scoring Unknown as a negative value
3.2	Are there available tools that can accurately and sufficiently identify the format of a file encoded in this format?	2 = Yes, 0 = N/A, -2 = No or Unknown	For most formats, an internal signature for the exact version in PRONOM is preferable	NEW QUESTION
3.3	Is there software supported by current computing environments that can create the format?	1 = Yes, -1 = No or Unknown	For software and code, the "software" in this question is referring to the compiler	Rephrased question for clarity, removed N/A option, scoring Unknown as a negative value
3.4	Does the format require the use of compression?	-1 = Yes or Unknown, 1 = No	If the format is a wrapper that is uncompressed but can contain compressed bitstreams, the answer here would be 1 = No	Removed N/A option, scoring Unknown as a negative value
3.5	*If the format requires the use of compression*, can it be lossy?	-1 = Yes or Unknown, 0= N/A or No	If answer to previous question is No, N/A If answer to previous question is Unknown, Unknown	NEW QUESTION

3.6	Does the format support the embedding of additional content in other unrelated formats?	-1 = Yes or Unknown, 1 = No		NEW QUESTION Note for reviewers: We would especially like feedback on this question. As currently written, it's quite broad and could apply to almost any complex file format from PDF to PPTX to media wrappers to archive files. We would like to have a question that captures how some formats are inherently less transparent because they may contain other arbitrary formats embedded inside of them, which cannot be discerned simply by knowing what format a file is. **Is there a better way to phrase this question?**
	Total Score:	Highest possible: 6 (previously 7) Lowest possible: -7 (previously -7)		

Questions removed from Transparency:
- Does the available specification provide enough detail to allow basic analysis of the file format with hex editors or other inspection tools? (replaced with 3.2)
- Does the format rely on standard character or other encoding methods such as IEEE notations? (there is not enough data that can be consistently used to answer this question - over half of all responses were "Unknown")
- Is the source code of the software used to create the format available for little or no cost? (some formats can be created using multiple software packages, "little cost" is too vague)
- Does the format support user-definable compression levels or other quality settings that affect essential characteristics? (the answer to this question isn't necessarily applicable to all file created in that format)

Self-Documentation

	Question	Values	Scoring Guidance	Updates
4.1	Does the format natively include sufficient metadata to document the technical environment used to create (and/or modify) the file?	1 = Yes, -1 = No or Unknown		NEW QUESTION
4.2	Does the format support adding embedded descriptive and administrative metadata through manual or automated processes?	1 = Yes, -1 = No or Unknown		NEW QUESTION
4.3	Does the format support metadata that is robust enough for an accurate file analysis?	2 = Yes, -1 = No or Unknown		Rephrased question for clarity
	Total Score:	Highest possible: 4 (previously 7) Lowest possible: -3 (previously -5)		

Questions removed from Self-Documentation:
- Does the format support self-contained technical metadata? (rephrased into 4.1)
- Does the format support self-contained (embedded) descriptive metadata? (condensed and rephrased into 4.2)
- Does the format support self-contained administrative metadata? (condensed and rephrased into 4.2)
- Does the format metadata adhere to an international standard? (there is not enough data that can be consistently used to answer this question)

External Hardware Dependencies

	Question	Values	Scoring Guidance	Updates
5.1	Does the format require a specific hardware environment, such as a specific graphics card, chipset, or memory requirements, to process or interact with it?	-4 = Yes or Unknown, 2 = No		Scoring Unknown as a negative value
5.2	Does the format require specific playback hardware (e.g., Blu-Ray, Audio CD, etc) to transfer the format to the NARA environment?	-4 = Yes or Unknown, 2 = No		Scoring Unknown as a negative value
	Total Score:	Highest possible: 4 (previously 4) Lowest possible: -8 (previously -8)		

Questions removed from External Hardware Dependencies:
None

External Software Dependencies

	Question	Values	Scoring Guidance	Updates
6.1	Are renderers available?	2 = Yes, 0 = N/A, -2 = No or Unknown	N/A applies to character encodings or software applications/executables.	NEW QUESTION
6.2	Does the format rely on proprietary software, plug-ins, and/or scripts, etc. to render or execute files?	-2 = Yes or Unknown, 1 = No		Added "proprietary software", removed N/A, scoring Unknown as a negative value
6.3	Can the format be rendered or executed in more than one computing operating system?	2 = Yes, -2 = No or Unknown		Rephrased for clarity, removed N/A, scoring Unknown as a negative value
	Total Score:	Highest possible: 5 (previously 4) Lowest possible: -6 (previously -8)		

Questions removed from External Software Dependencies:
- Does the format rely on proprietary software to render or view files? (merged into 6.2)
- Does the format rely on special tools to render or view files? (overlap with 6.2)

Impact of Patents

	Question	Values	Scoring Guidance	Updates
7.1	Is the format subject to patent claims that may impede the development of open source tools for opening and managing the files?	-2 = Yes or Unknown, 2 = No		Removed N/A, scoring Unknown as a negative value, giving heavier weight to both positive and negative scores
7.2	Does the format have open source license terms?	2 = Yes, -2 = No or Unknown		Removed N/A, scoring Unknown as a negative value, giving heavier weight to both positive and negative scores
	Total Score:	Highest possible: 4 (previously 4) Lowest possible: -4 (previously -4)		

Questions removed from Impact of Patents:
- Have the patent claims expired? (there is not enough data that can be consistently used to answer this question - over half of all responses were "Unknown")
- Are there fees associated with the format as a result of patent claims? (there is not enough data that can be consistently used to answer this question)

Technical Protection Mechanisms

	Question	Values	Scoring Guidance	Updates
8.1	Does the format require the use of encryption?	-4 = Yes or Unknown, 2 = No		Removed N/A, scoring Unknown as a negative value, giving heavier weight to Yes
8.2	***If the format requires the use of encryption***, is it robust?	-1 = Yes or Unknown, 0 = N/A or No	If answer to previous question is No, N/A If answer to previous question is Unknown, Unknown	Added conditional IF statement, scoring Unknown as a negative value
8.3	Does the format natively allow the use of technical protection measures (e.g. digital rights management)?	-2 = Yes or Unknown, 2 = No		Rephrased for clarity
8.4	Does the format natively allow for embedded information such as watermarking?	-1 = Yes or Unknown, 1 = No		Removed N/A, scoring Unknown as a negative value
	Total Score:	Highest possible: 5 (previously 10) Lowest possible: -8 (previously -6)		

Questions removed from Technical Protection Mechanisms:
- Does the format have capability to encrypt all or part of the resulting file? (we care less about a format's ability to encrypt than whether it is required, because any file can be encrypted after the fact)

Format Age

	Question	Values	Scoring Guidance	Updates
9.1	When was the format specification for this version first created?	4 digit year		Added "for this version" for clarity
9.2	Risk Factor (created)	0 = 5 years old or less, -2 = 6-15 years, -4 = 16+ years or Unknown		None
9.3	When was the format specification for this version last updated?	4 digit year		Added "for this version" for clarity
9.4	Risk Factor (last updated)	0=5 years old or less, -2 for 6-15 years, -4 for 16+ years		None
	Total Score:	Highest possible: 0 (previously 0) Lowest possible: -8 (previously -8)		

Questions removed from Format Age:
None

Risk/Sustainability Factor

	Question	Values	Scoring Guidance	Updates
10.1	Total Format Risk/Sustainability Factor Numeric Score	Highest possible: 42 Lowest possible: -49	Sum of all preceding category total scores	Previous highest possible: 56 Previous lowest possible: -58
10.2	Risk/Sustainability Factor Status	Low = 11 to 42 Moderate = -21 to 10 High= -49 to -22		Previous Low: 22 to 56 Previous Moderate: -12 to 23 Previous High: -58 to -13 Note to reviewers: These are the ranges that achieved roughly the same distribution of Low/Moderate/High risk levels in a random 10% sample. We will likely adjust this slightly as we rescore the whole matrix.

Questions removed from Risk/Sustainability Factor:
None

Prioritization

	Question	Values	Scoring Guidance	Updates
11.1	Number of identified files in the combined NARA holdings	Number from holdings profile		None Note to reviewers: This column is not included in the publicly-shared version of the Risk Matrix.
11.2	Percentage of 2 billion files in NARA ERA holdings			None
11.3	Prevalence: Format Adoption Level as measured by Proportion of File Format in the Overall 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%		None
11.4	Feasibility: Ability to Convert (tools exist for conversion that does not alter content in unacceptable ways; NARA can perform acceptable transformations)	No acceptable tools available in the marketplace = -5 Acceptable tools exist but NARA does not have them = -3 No transformation is needed = 0 Transformation already performed at NARA = 3 Preferred/Acceptable Format as per Transfer Guidance = 5		Scoring "No transformation is needed" as 0 rather than 5
11.5	NARA Total	Total Format Risk/Sustainability Factor		None

		Numeric Score + Prevalence + Feasibility		

Questions removed from Prioritization:
None