

2024 NARA Risk Matrix Updates

NARA uses the Risk 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 performed a comprehensive review and update of all the questions and categories of questions. We made 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. These changes were available for public comment in May 2024. After reviewing and incorporating feedback, the updated version of the Risk Matrix was released in September 2024, encompassing the updates below.

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 were removed.

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	<i>If there is a published specification</i> , has it been approved and published by an internationally recognized standards organization?	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	When was the format specification for this version last updated (or, if the last updated date is not available, when was it created)?	0 = 5 years old or less, -2 = 6-15 years, -4 = 16+ years or Unknown		Moved question into this section from age-related risk factors, added “for this version” for clarity
	Total Score:	Highest possible: 6 (previously 10) Lowest possible: -7 (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? (there is not enough data that can be consistently used to answer this question)
- 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 more than 2,000 files in 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 or 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	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 to avoid for creation and transfer of permanent materials?	1 = No, -2 = Yes or Unknown		Reworded this question from “Have the archives or library communities identified the format as one they prefer for creation and transfer of permanent materials?” Removed N/A option, scoring Unknown as a negative value
	Total Score:	Highest possible: 5 (previously 10) Lowest possible: -3 (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 character-based, rather than binary?	1 = Yes, 0 = N/A, -1 = No or Unknown	N/A applies to formats that can be either character-based or binary	Reworded this question from "Is the format human readable and can it be opened with a text editor?" Removed N/A option, scoring Unknown as a negative value
3.2	Is there an internal signature in an authoritative format registry that can be used to identify a file in this format?	2 = Yes, 0 = N/A, -2 = No or Unknown	N/A applies to character encodings or software applications/executables	NEW QUESTION
3.3	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
3.4	<i>If the format requires the use of compression</i> , is it required to 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
3.5	Does this format support declared or hidden embedded data?	-1 = Yes or Unknown, 0 = N/A, 1 = No	If the format is compound (e.g. ESRI Shapefile) or an AV container, the answer here is N/A	NEW QUESTION

3.6	<i>If the format supports declared or hidden embedded data</i> , does the format list or provide standard information about embedded data?	1 = Yes, 0 = N/A, -1 = No or Unknown	If answer to previous question is No, N/A If answer to previous question is Unknown, Unknown	NEW QUESTION
	Total Score:	Highest possible: 5 (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)
- Is the software used to create the format supported by current computing environments? (replaced with similar question 6.5)
- 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 metadata that documents the file and the technical environment used to create (and/or modify) it?	2 = Yes, -2 = 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
	Total Score:	Highest possible: 3 (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?	-6 = Yes or Unknown, 2 = No		Scoring Unknown as a negative value, giving heavier weight to Yes or Unknown
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: -10 (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, -4 = No or Unknown	N/A may apply to character encodings or software applications/executables.	NEW QUESTION
6.2	If renderers are available , is at least one of them open source?	1 = Yes, 0 = N/A, -1 = No or Unknown	If answer to previous question is No, N/A If answer to previous question is Unknown, N/A	NEW QUESTION
6.3	Does the format rely on plug-ins and/or scripts, etc. to render or execute files?	-4 = Yes or Unknown, 0 = N/A, 2 = No	N/A may apply to character encodings or software applications/executables.	Scoring Unknown as a negative value, giving heavier weight to Yes or Unknown
6.4	Can the format be rendered or executed in more than one computing operating system?	2 = Yes, 0 = N/A, -2 = No or Unknown		Rephrased for clarity, scoring Unknown as a negative value, giving heavier weight to No or Unknown
6.5	Is there software supported by current computing environments that can create the format?	1 = Yes, 0 = N/A, -1 = No or Unknown	For software and code, the “software” in this question is referring to the compiler	Rephrased question for clarity, scoring Unknown as a negative value
	Total Score:	Highest possible: 8 (previously 4) Lowest possible: -11 (previously -8)		

Questions removed from External Software Dependencies:

- Does the format rely on proprietary software to render or view files? (merged into 6.3)

- Does the format rely on special tools to render or view files? (overlap with 6.3)

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	<i>If the format requires the use of encryption</i> , is it robust (cryptographically secure)?	-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
	Total Score:	Highest possible: 4 (previously 10) Lowest possible: -7 (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)

Risk/Sustainability Factor

	Question	Values	Scoring Guidance	Updates
	Total Format Risk/Sustainability Factor Numeric Score	Highest possible: 39 Lowest possible: -52	Sum of all preceding category total scores	Previous highest possible: 56 Previous lowest possible: -58
	Risk/Sustainability Factor Status	Low = 23 to 39 Moderate = -22 to 22 High= -52 to -23		Previous Low: 22 to 56 Previous Moderate: -12 to 23 Previous High: -58 to -13

Questions removed from Risk/Sustainability Factor:

None

Prioritization

These questions do not count towards the Risk/Sustainability Factor. They are used to calculate the NARA Total, factoring in both risk and prioritization.

	Question	Values	Scoring Guidance	Updates
	Percentage of 2 billion files in NARA ERA holdings			None
	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
	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
	NARA Total	Total Format Risk/Sustainability Factor Numeric Score + Prevalence + Feasibility		None

Questions removed from Prioritization:
None