Preservation Action Plan: Textual/Word Processing Records
National Archives and Records Administration (NARA)

DRAFT 20190801
Template: 201907

Electronic Record or Digital Surrogate Types and Associated Formats

Records created through the use of a word processing or text processing tool. Examples include but are not limited to: Minutes of meetings, organizational charts, diaries, calendars, correspondence, reports, briefing books, legal opinions, and directives.

Essential Characteristics of this Record Type

The rapid evolution of textual data types, and the variety of agency usage, limits defining characteristics only in the manner in which NARA can currently accept the transfer of textual files: essential characteristics are currently dependent on the export format. While embedded or hyperlinked content may be found across record types, we may not know much about its structure, data type, or record type. We cannot capture every possible structural nuance of the records.

Some appearance characteristics are discussed in the Structure Characteristics section, since the appearance of a textual record can be partially based on technical structure that allows the record to be manipulated and displayed in a different manner. If the appearance is fixed, as for official publication, then the dimensions must be maintained. If the appearance is determined by user preference or is otherwise mutable, then the dimensions are probably not as important.

Questions which must be considered:
- Would a change in the record's appearance alter its meaning?
- Does changing the record's appearance diminish its value? For example, if the records have been appraised as permanent for their informational value, and not evidential, then appearance characteristics may not need to be preserved.
- Would a change in the record's technical structure alter its appearance?
- Would a change in the record's technical structure affect its possible behaviors?

Built in tools such as macros with built in reports or external links are related to more record or data types which have their own essential characteristics.

Appearance

Name	Definition	Function Description
Language	A language identifier specifying the natural language(s) used in the document	
Fonts	A list of fonts used in the document, and an indication of whether or not a font is embedded in a document. ● Typeface (Arial, Times New Roman, etc.) ● Size (10 pt, 18 pt, etc.) ● Pitch ● Spacing ● Emphasis (bold, italic, strikethrough, underline, etc.)	Stores the names of all fonts (embedded and non-embedded) used in a document to assist in rendering the content and identifying the documents with potential long-term preservation risks. There will always be a font in order to have text, however, there are other variables that determine whether a distinct font is a core characteristic. For example, where appearance is fixed, such as in a publication, then font is core.
Size	Document "page" size(s), e.g., Letter, Legal, A4	Physical or virtual dimensions. These characteristics are conditional. If the appearance is important, such as a publication, then this must be maintained. However, if only the content ("words absent their formatting") of the document is important, such as meeting minutes, then these dimensional characteristics are not core.
Layout Orientation	Orientation of the "pages," e.g., portrait versus landscape.	Needed for understanding content and context of record.
Text Orientation	Orientation of the text, e.g., horizontal, vertical, custom axes	Needed for understanding content and context of record.
Color	Identification of the use of color in text and layout elements, e.g. borders, boxes.	Color is essential if it bears meaning and/or value, or if the appearance is fixed.

	● Hue: color family or name ● Saturation: purity or sharpness ● Brightness: shade or tint ● Contrast: range of optical density or tone	
Pagination	Pagination and content sequence in the document.	If present, then needed for contextual understanding of the record.
Forms	Form elements embedded in the document	
Annotations	Annotations included in the document.	
Tables	Includes content structured in a table feature, such as a dynamically generated table of contents.	
Graphics	Images embedded in the document, or graphics created using a word processing graphics feature.	
Outline	Includes content in an outline structure.	
Transparency	The use of transparency features for graphics or text in the document.	
Layers	The presence of layered graphic elements in the document.	

Structure

Name	Definition	Function Description
Schema	- Lists - Arrays - Models - Trees - Tables - Linkage: connection between or within records. If connections exist, then they are core - Template. Although there is a generally agreed upon composition that exists for written textual records (business and government standards) such as with a letter, the placement of addresses, date, salutation, signature, content, enclosure, carbon copies, if there is the existence of a template needed to generate a textual record, then the template is core.	
Character Encoding	Encoding schema, e.g., US-ASCII, EBCDIC, UTF-8.	Required for the proper parsing and rendering of the record content.
Word Count	Total number of words in the document	Valuable for evaluating the completeness of the content after transformations, but can be highly variable between tools.
Character Count	Total number of characters in the document	Valuable for evaluating the completeness of the content after transformations, but can be highly variable between tools.

Line Count	Total number of lines in the document	Valuable for evaluating the completeness of the content after transformations, but can be highly variable between tools.
Paragraph Count	Total number of paragraphs in the document	Valuable for evaluating the completeness of the content after transformations, but can be highly variable between tools.
Table Count	Total number of tables in the document	Valuable for evaluating the completeness of the content after transformations, but can be highly variable between tools.
Graphics Count	Total number of graphics in the document	Valuable for evaluating the completeness of the content after transformations, but can be highly variable between tools.

Behavior

Name	Definition	Function Description
Macros and Linkages	User-created automated functions; User-defined linkages within and between documents/files.	Essential characteristics are currently dependent on the export format.

Context

Name	Definition	Function Description
Series	Identification of the document's inclusion in a series, and enumeration if present.	

Descriptive Metadata	Descriptive metadata embedded in the record file header.	Information contained within the record (intrinsic) that refers to the intellectual content of material and aids discovery of such materials, for example, Caption/Subject/Date/Event/Transaction can all add value to the record.

Current NARA Transfer Guidance for this Record Type

Preferred: OpenDocument Text Format (ODF) 1.0, PDF/A-1, PDF/A-2

Acceptable: PDF 1.0-1.7, Microsoft Word Office Open XML (DOCX), Microsoft Word 97 Binary Document Format (DOC)

Current NARA Public Access/Reference Format(s) for this Record Type
This Plan references existing public access file formats for electronic records at NARA, determined with a survey of the available public access formats in the National Archives Catalog. These references do not represent recommended public access formats under NARA policies. They are intended for informational purposes only.

Reference Format: PDF

Public Access Format: PDF

WordPerfect versions 6-12 (All Platforms)

NARA Format ID: NF00452

Extension(s)
- wpd
- wp6
- wp7

Documentation
- https://en.wikipedia.org/wiki/WordPerfect
- http://fileformats.archiveteam.org/wiki/WordPerfect

Risk and Prioritization Analysis

☐ **Low Risk**

✓ Moderate Risk

☐ **High Risk**

1   Numeric Risk Rating

-1   Numeric Prioritization Rating

Proposed Preservation Plan

☐ **Retain** file format in its existing format.

✓ Transform file to a new format.

      Selected Format: ODF 1.0

☐ **Procure/develop tools** to preserve, manage and provide access to records of this type in their existing form.

✓ Procure/develop tools to transform the format to the preferred normalized form.

☐ **Provide Additional Information** so that the record type remains understandable/usable over time.

☐ **Explore Additional Options**

Preferred Processing and Transformation Tool(s)

- Universal Office Converter - Commandline library that converts between any document format supported by LibreOffice/OpenOffice (https://github.com/dagwieers/unoconv)

Preferred Viewer/Access Software

- Libre Office
- OpenOffice

- Adobe Acrobat for PDF

Rich Text Format versions 1, 1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9

NARA Format ID: NF00398

Extension(s):
- rtf

Documentation

- https://en.wikipedia.org/wiki/Rich_Text_Format
- https://www.loc.gov/preservation/digital/formats/fdd/fdd000473.shtml
- http://fileformats.archiveteam.org/wiki/RTF

Risk and Prioritization Analysis

☐ **Low Risk**

✓ **Moderate Risk**

☐ **High Risk**

5 Numeric Risk Rating

1 Numeric Prioritization Rating

Proposed Preservation Plan

☐ **Retain** file format in its existing format.

✓ Transform file to a new format.
> **Selected Format:** ODF 1.0

☐ **Procure/develop tools** to preserve, manage and provide access to records of this type in their existing form.

✓ Procure/develop tools to transform the format to the preferred normalized form.

☐ **Provide Additional Information** so that the record type remains understandable/usable over time.

☐ **Explore Additional Options**

Preferred Processing and Transformation Tool(s)

- Universal Office Converter - Commandline library that converts between any document format supported by LibreOffice/OpenOffice (https://github.com/dagwieers/unoconv)

Preferred Viewer/Access Software

- Libre Office
- OpenOffice

- Adobe Acrobat for PDF

WordPerfect for DOS 5.0, 5.1

NARA Format ID:
- **NF00455 (5.0)**
- **NF00453 (5.1)**
- **NF00456 (Secondary Files)**

Extension(s):
- wp
- wp5
- wpd (Secondary Files)

Documentation
- https://en.wikipedia.org/wiki/WordPerfect
- http://fileformats.archiveteam.org/wiki/WordPerfect
- https://winworldpc.com/product/wordperfect/50-dos
- https://winworldpc.com/product/wordperfect/5x-dos

Risk and Prioritization Analysis

☐ **Low Risk**

✓ **Moderate Risk**

☐ **High Risk**

1   Numeric Risk Rating

-3 Numeric Prioritization Rating (The lower the number, the higher the priority)

Proposed Preservation Plan

☐ **Retain** file format in its existing format.

✓ Transform file to a new format.
   Selected Format: PDF

☐ **Procure/develop tools** to preserve, manage and provide access to records of this type in their existing form.

✓ Procure/develop tools to transform the format to the preferred normalized form.

☐ **Provide Additional Information** so that the record type remains understandable/usable over time.

☐ **Explore Additional Options**

Preferred Processing and Transformation Tool(s)

- Universal Office Converter - Commandline library that converts between any document format supported by LibreOffice/OpenOffice (https://github.com/dagwieers/unoconv)

Preferred Viewer/Access Software

- Libre Office
- OpenOffice
- Adobe Acrobat for PDF

WordPerfect for Windows 5.1, 5.2

NARA Format ID:
- **NF00453 (5.1)**
- **NF00454 (5.2)**
- **NF00456 (Secondary Files)**

Extension(s):
- wp
- wp5
- wpd (Secondary Files)

Documentation
- https://en.wikipedia.org/wiki/WordPerfect
- http://fileformats.archiveteam.org/wiki/WordPerfect
- https://winworldpc.com/product/wordperfect/5x-win

Risk and Prioritization Analysis

☐ **Low Risk**

✓ Moderate Risk

☐ **High Risk**
1 Numeric Risk Rating
-3 Numeric Prioritization Rating (The lower the number, the higher the priority)

Proposed Preservation Plan

☐ **Retain** file format in its existing format.

✓ Transform file to a new format.
    Selected Format: PDF
☐ **Procure/develop tools** to preserve, manage and provide access to records of this type in their existing form.

✓ Procure/develop tools to transform the format to the preferred normalized form.
☐ **Provide Additional Information** so that the record type remains understandable/usable over time.
☐ **Explore Additional Options**

Preferred Processing and Transformation Tool(s)

- Universal Office Converter - Commandline library that converts between any document format supported by LibreOffice/OpenOffice (https://github.com/dagwieers/unoconv)

Preferred Viewer/Access Software

- Libre Office
- OpenOffice
- Adobe Acrobat for PDF

Microsoft Word for MS-DOS version 3

NARA Format ID: NF00308

Extension(s):
- doc

Documentation
- http://www.forensicswiki.org/wiki/Word_Document_%28DOC%29
- https://msdn.microsoft.com/en-us/library/office/cc313153(v=office.12).aspx
- https://en.wikipedia.org/wiki/History_of_Microsoft_Word
- https://en.wikipedia.org/wiki/Microsoft_Word#Release_history
- https://en.wikipedia.org/wiki/Microsoft_Office_3.0

Risk and Prioritization Analysis

☐ **Low Risk**

✓ **Moderate Risk**

☐ **High Risk**

9 **Numeric Risk Rating**

9 **Numeric Prioritization Rating**

Proposed Preservation Plan

☐ **Retain** file format in its existing format.

✓ **Transform** file to a new format.
> **Selected Format:** PDF 1.0

☐ **Procure/develop tools** to preserve, manage and provide access to records of this type in their existing form.

✓ **Procure/develop tools** to transform the format to the preferred normalized form.

☐ **Provide Additional Information** so that the record type remains understandable/usable over time.

☐ **Explore Additional Options**

Justification: Although an open specification is available, the format is proprietary and subject to patents and copyright. This version is no longer supported.

Preferred Processing and Transformation Tool(s)

- Universal Office Converter - Command line library that converts between any document format supported by LibreOffice/OpenOffice (https://github.com/dagwieers/unoconv)

Preferred Viewer/Access Software

- Libre Office
- OpenOffice
- Adobe Acrobat for PDF

Microsoft Word for MS-DOS versions 4, 5, 5.5

NARA Format ID: NF00309

Extension(s):
- doc

Documentation
- http://www.forensicswiki.org/wiki/Word_Document_%28DOC%29
- https://msdn.microsoft.com/en-us/library/office/cc313153(v=office.12).aspx
- https://en.wikipedia.org/wiki/History_of_Microsoft_Word
- https://en.wikipedia.org/wiki/Microsoft_Word#Release_history
- https://en.wikipedia.org/wiki/Microsoft_Office_3.0

Risk and Prioritization Analysis

☐ **Low Risk**

✓ **Moderate Risk**

☐ **High Risk**

7 **Numeric Risk Rating**

7 **Numeric Prioritization Rating**

Proposed Preservation Plan

☐ **Retain** file format in its existing format.

✓ Transform file to a new format.
  Selected Format: PDF 1.0

☐ **Procure/develop tools** to preserve, manage and provide access to records of this type in their existing form.

✓ Procure/develop tools to transform the format to the preferred normalized form.

☐ **Provide Additional Information** so that the record type remains understandable/usable over time.

☐ **Explore Additional Options**

Justification: Although an open specification is available, the format is proprietary and subject to patents and copyright. This version is no longer supported.

Preferred Processing and Transformation Tool(s)

- Universal Office Converter - Command line library that converts between any document format supported by LibreOffice/OpenOffice (https://github.com/dagwieers/unoconv)

Preferred Viewer/Access Software

- Libre Office
- OpenOffice
- Adobe Acrobat for PDF

Microsoft Word for Macintosh version 6.0

Extension(s):
- doc

Documentation:

- https://en.wikipedia.org/wiki/History_of_Microsoft_Word#Word_2001/Word_X
- https://msdn.microsoft.com/en-us/library/cc313153.aspx
- http://lowendmac.com/2013/microsoft-word-for-mac-faq/

Risk and Prioritization Analysis

☐ **Low Risk**

✓ **Moderate Risk**

☐ **High Risk**

8 Numeric Risk Rating

6 Numeric Prioritization Rating

Proposed Preservation Plan

☐ **Retain** file format in its existing format.

✓ **Transform** file to a new format.
> **Selected Format:** PDF 1.0

☐ **Procure/develop tools** to preserve, manage and provide access to records of this type in their existing form.

✓ **Procure/develop tools** to transform the format to the preferred normalized form.

☐ **Provide Additional Information** so that the record type remains understandable/usable over time.

☐ **Explore Additional Options**

Justification: Although an open specification is available, the format is proprietary and subject to patents and copyright. This version is no longer supported.

Preferred Processing and Transformation Tool(s)

- Universal Office Converter - Command line library that converts between any document format supported by LibreOffice/OpenOffice (https://github.com/dagwieers/unoconv)

Preferred Viewer/Access Software

- Adobe Acrobat for PDF

Microsoft Word for Macintosh version X

Extension(s):
- doc

Documentation:

- https://en.wikipedia.org/wiki/Microsoft_Word
- https://en.wikipedia.org/wiki/Microsoft_Office_2001

Risk and Prioritization Analysis

☐ **Low Risk**

✓ **Moderate Risk**

☐ **High Risk**

7   Numeric Risk Rating

5   Numeric Prioritization Rating

Proposed Preservation Plan

☐ **Retain** file format in its existing format.

✓ Transform file to a new format.
   > **Selected Format:** PDF 1.0

☐ **Procure/develop tools** to preserve, manage and provide access to records of this type in their existing form.

✓ Procure/develop tools to transform the format to the preferred normalized form.

☐ **Provide Additional Information** so that the record type remains understandable/usable over time.

☐ **Explore Additional Options**

Justification: This is a proprietary format, subject to patents and copyright. This version is no longer supported.

Preferred Processing and Transformation Tool(s)

- Universal Office Converter - Command line library that converts between any document format supported by LibreOffice/OpenOffice (https://github.com/dagwieers/unoconv)

Preferred Viewer/Access Software

- Adobe Acrobat for PDF

Microsoft Word for Windows 97-2003

Extension(s)
- doc
- word-97
- dot (Template)

Documentation:
- https://en.wikipedia.org/wiki/Microsoft_Word
- https://en.wikipedia.org/wiki/Microsoft_Office_97
- http://home.bt.com/tech-gadgets/computing/software/is-it-safe-to-use-an-old-version-of-microsoft-office-11364006850132

Risk and Prioritization Analysis

☐ **Low Risk**

✓ **Moderate Risk**

☐ **High Risk**

16 Numeric Risk Rating

16 Numeric Prioritization Rating

Proposed Preservation Plan

✓ **Retain** file format in its existing format.

Transform file to a new format.

☐ **Procure/develop tools** to preserve, manage and provide access to records of this type in their existing form.

Procure/develop tools to transform the format to the preferred normalized form.

☐ **Provide Additional Information** so that the record type remains understandable/usable over time.

☐ **Explore Additional Options**

Justification: This format has been listed as "acceptable" by NARA.

Preferred Processing and Transformation Tool(s)

- Universal Office Converter - Command line library that converts between any document format supported by LibreOffice/OpenOffice (https://github.com/dagwieers/unoconv)

Preferred Viewer/Access Software

- Microsoft Office - Word

Microsoft Word for Windows 6.0, 95

Extension(s)
- doc
- dot (Template)

Documentation:
- https://en.wikipedia.org/wiki/Microsoft_Word
- https://en.wikipedia.org/wiki/History_of_Microsoft_Word
- https://en.wikipedia.org/wiki/Microsoft_Office_95
- http://home.bt.com/tech-gadgets/computing/software/is-it-safe-to-use-an-old-version-of-microsoft-office-11364006850132

Risk and Prioritization Analysis

☐ **Low Risk**

✓ **Moderate Risk**

☐ **High Risk**

14 Numeric Risk Rating

14 Numeric Prioritization Rating

Proposed Preservation Plan

✓ **Retain** file format in its existing format.

Transform file to a new format.

☐ **Procure/develop tools** to preserve, manage and provide access to records of this type in their existing form.

Procure/develop tools to transform the format to the preferred normalized form.

☐ **Provide Additional Information** so that the record type remains understandable/usable over time.

☐ **Explore Additional Options**

Justification: This format has been listed as "acceptable" by NARA.

Preferred Processing and Transformation Tool(s)

- Universal Office Converter - Command line library that converts between any document format supported by LibreOffice/OpenOffice (https://github.com/dagwieers/unoconv)

Preferred Viewer/Access Software

- Microsoft Office - Word

Microsoft Word Office Open XML

Extension(s):
- docx
- docm
- dotx (Template
- dotm (Template)

Documentation:
- https://en.wikipedia.org/wiki/Office_Open_XML
- https://en.wikipedia.org/wiki/Standardization_of_Office_Open_XML
- https://en.wikipedia.org/wiki/Microsoft_Open_Specification_Promise
- https://en.wikipedia.org/wiki/Microsoft_Office_XML_formats
- https://docs.microsoft.com/en-us/office/open-xml/understanding-the-open-xml-file-formats
- https://msdn.microsoft.com/en-us/library/dd773189(v=office.12).aspx

Risk and Prioritization Analysis

✓ **Low Risk**

☐ **Moderate Risk**

☐ **High Risk**

29 Numeric Risk Rating

29 Numeric Prioritization Rating (The lower the number, the higher the priority)

Proposed Preservation Plan

✓ **Retain** file format in its existing format.

☐ **Transform** file to a new format.

☐ **Procure/develop tools** to preserve, manage and provide access to records of this type in their existing form.

☐ **Procure/develop tools** to transform the format to the preferred normalized form.

☐ **Provide Additional Information** so that the record type remains understandable/usable over time.

☐ **Explore Additional Options**

Justification: This format is considered acceptable by NARA.

Preferred Processing and Transformation Tool(s)

- Universal Office Converter - Command line library that converts between any document format supported by LibreOffice/OpenOffice (https://github.com/dagwieers/unoconv)

Preferred Viewer/Access Software

- Microsoft Office - Word

Write for Windows Document 3.0

Extension(s)
- wri

Documentation:

- https://en.wikipedia.org/wiki/Microsoft_Write
- https://fileinfo.com/extension/wri

Risk and Prioritization Analysis

☐ **Low Risk**

✓ **Moderate Risk**

☐ **High Risk**

1 Numeric Risk Rating

-4 Numeric Prioritization Rating

Proposed Preservation Plan

☐ **Retain** file format in its existing format.

✓ Transform file to a new format.

      Transform into PDF

☐ **Procure/develop tools** to preserve, manage and provide access to records of this type in their existing form.

✓ Procure/develop tools to transform the format to the preferred normalized form.

☐ **Provide Additional Information** so that the record type remains understandable/usable over time.

☐ **Explore Additional Options**

Justification: This is an outdated, not supported, proprietary format.

Preferred Processing and Transformation Tool(s)

- Universal Office Converter - Command line library that converts between any document format supported by LibreOffice/OpenOffice (https://github.com/dagwieers/unoconv)
- File converter online: https://wri-to-pdf.file-converter-online.com/
- WRI to PDF API: https://www.convertapi.com/wri-to-pdf

Preferred Viewer/Access Software

- Adobe Acrobat for PDF

Microsoft Word for Macintosh versions 5.0, 5.2

Extension(s):
- doc

Documentation:

- https://docs.microsoft.com/en-us/deployoffice/compat/office-file-format-reference
- http://www.nationalarchives.gov.uk/PRONOM/Format/proFormatSearch.aspx?status=detailReport&id=407
- https://msdn.microsoft.com/en-us/library/cc313153.aspx
- https://msdn.microsoft.com/en-us/openspecifications/dn750984
- https://stackoverflow.com/questions/4346117/how-can-i-view-open-a-word-document-in-my-browser-using-with-php-or-html
- https://mspoweruser.com/microsoft-releases-the-source-code-of-ms-dos-on-github/
- http://lowendmac.com/2013/microsoft-word-for-mac-faq/

Risk and Prioritization Analysis

☐ **Low Risk**

✓ **Moderate Risk**

☐ **High Risk**

2   Numeric Risk Rating

0   Numeric Prioritization Rating

Justification: Of the Word for Mac formats, this is considered to be the superior version, and was in wide circulation among Mac users. The format is proprietary, outdated, and no longer supported.

Proposed Preservation Plan

☐ **Retain** file format in its existing format.

✓ Transform file to a new format.

> Transform into PDF

☐ **Procure/develop tools** to preserve, manage and provide access to records of this type in their existing form.

✓ Procure/develop tools to transform the format to the preferred normalized form.

☐ **Provide Additional Information** so that the record type remains understandable/usable over time.

☐ **Explore Additional Options**

Preferred Processing and Transformation Tool(s)

- Universal Office Converter - Command line library that converts between any document format supported by LibreOffice/OpenOffice (https://github.com/dagwieers/unoconv)

Preferred Viewer/Access Software

- Adobe Acrobat for PDF

DEC WPS Plus

Extension(s)
- dx

Documentation:

- https://fileinfo.com/extension/dx
- https://www.wikidata.org/wiki/Q51994258
- http://www.nationalarchives.gov.uk/PRONOM/Format/proFormatSearch.aspx?status=detailReport&id=438

Risk and Prioritization Analysis

☐ **Low Risk**

✓ **Moderate Risk**

☐ **High Risk**

11 Numeric Risk Rating

1 Numeric Prioritization Rating

Proposed Preservation Plan

☐ **Retain** file format in its existing format.

✓ Transform file to a new format.
> TBD, Preferably PDF

☐ **Procure/develop tools** to preserve, manage and provide access to records of this type in their existing form.

✓ Procure/develop tools to transform the format to the preferred normalized form.

☐ **Provide Additional Information** so that the record type remains understandable/usable over time.

☐ **Explore Additional Options**

Justification: Very rare, outdated and unsupported file.

Preferred Processing and Transformation Tool(s)

- TBD

Preferred Viewer/Access Software

- Adobe Acrobat for PDF

ASCII 7-bit Text

NARA Format ID: NF00113

Extension(s):
- txt
- asc
- csv
- tab

Documentation:

- http://www.neurophys.wisc.edu/comp/docs/ascii/
- ASCII and 7-bit ASCII are synonymous

Risk and Prioritization Analysis

✓ **Low Risk**

☐ **Moderate Risk**

☐ **High Risk**

37 Numeric Risk Rating

36 Numeric Prioritization Rating

Proposed Preservation Plan

✓ **Retain** file format in its existing format.

☐ **Transform** file to a new format.

☐ **Procure/develop tools** to preserve, manage and provide access to records of this type in their existing form.

☐ **Procure/develop tools** to transform the format to the preferred normalized form.

☐ **Provide Additional Information** so that the record type remains understandable/usable over time.

☐ **Explore Additional Options**

Justification: ASCII-7bit it a preferred format of the National Archives.

Preferred Processing and Transformation Tool(s)

- Universal Office Converter - Command line library that converts between any document format supported by LibreOffice/OpenOffice (https://github.com/dagwieers/unoconv)

Preferred Viewer/Access Software

- Notepad++ (https://notepad-plus-plus.org/)

ASCII 8-bit Text

NARA Format ID: NF00114

Extension(s):
- txt
- asc
- csv
- tab

Documentation:

- http://www.neurophys.wisc.edu/comp/docs/ascii/

Risk and Prioritization Analysis

✓ Low Risk
☐ Moderate Risk
☐ High Risk
40 Numeric Risk Rating
40 Numeric Prioritization Rating

Proposed Preservation Plan

✓ **Retain** file format in its existing format.
☐ **Transform** file to a new format.
☐ **Procure/develop tools** to preserve, manage and provide access to records of this type in their existing form.
☐ **Procure/develop tools** to transform the format to the preferred normalized form.
☐ **Provide Additional Information** so that the record type remains understandable/usable over time.
☐ **Explore Additional Options**

Justification: ASCII-7bit it a preferred format of the National Archives.

Preferred Processing and Transformation Tool(s)

- Universal Office Converter - Command line library that converts between any document format supported by LibreOffice/OpenOffice (https://github.com/dagwieers/unoconv)

Preferred Viewer/Access Software

- Notepad++ (https://notepad-plus-plus.org/)

Extended ASCII, Variable-width encoding Unicode (Universal Coded Character Set) Transformation Format (UTF)

NARA Format ID: NF00429

Extension(s):
- utf
- txt
- utf8
- utf16

Documentation:
- https://en.wikipedia.org/wiki/Comparison_of_Unicode_encodings
- https://en.wikipedia.org/wiki/UTF-16
- https://encoding.spec.whatwg.org/#security-background
- https://tools.ietf.org/html/rfc2781
- https://ipfs.io/ipfs/QmXoypizjW3WknFiJnKLwHCnL72vedxjQkDDP1mXWo6uco/wiki/UTF-9_and_UTF-18.html

Risk and Prioritization Analysis

✓ Low Risk

☐ Moderate Risk
☐ High Risk
44 Numeric Risk Rating
44 Numeric Prioritization Rating

Proposed Preservation Plan

✓ **Retain** file format in its existing format.
☐ **Transform** file to a new format.
☐ **Procure/develop tools** to preserve, manage and provide access to records of this type in their existing form.
☐ **Procure/develop tools** to transform the format to the preferred normalized form.
☐ **Provide Additional Information** so that the record type remains understandable/usable over time.
☐ **Explore Additional Options**

Justification: UTF-8 and UTF-16 are NARA preferred transfer formats.

Preferred Processing and Transformation Tool(s)

- Universal Office Converter - Command line library that converts between any document format supported by LibreOffice/OpenOffice (https://github.com/dagwieers/unoconv)

Preferred Viewer/Access Software

- Notepad++ (https://notepad-plus-plus.org/)

Plain Text

NARA Format ID: NF00360

Extension(s):
- Plain_Text
- txt

Documentation:

- https://en.wikipedia.org/wiki/Plain_text
- https://en.wikipedia.org/wiki/ISO/IEC_8859
- https://en.wikipedia.org/wiki/Text_file
- https://www.w3.org/Protocols/rfc1341/7_1_Text.html

Risk and Prioritization Analysis

☐ **Low Risk**

✓ Moderate Risk

☐ **High Risk**
-10 Numeric Risk Rating
-10 Numeric Prioritization Rating

Proposed Preservation Plan

- ✓ **Retain** file format in its existing format.
- ☐ **Transform** file to a new format.
- ☐ **Procure/develop tools** to preserve, manage and provide access to records of this type in their existing form.
- ☐ **Procure/develop tools** to transform the format to the preferred normalized form.
- ☐ **Provide Additional Information** so that the record type remains understandable/usable over time.
- ☐ **Explore Additional Options**

Justification: A text file is a type of container, and the contents may or may not be plain text (a type of content). The ".txt" file extension is associated with Microsoft Windows operating systems, but many other extensions can be used.

Preferred Processing and Transformation Tool(s)

- Universal Office Converter - Command line library that converts between any document format supported by LibreOffice/OpenOffice (https://github.com/dagwieers/unoconv)

Preferred Viewer/Access Software

- Adobe Acrobat for PDF

Microsoft Word for Windows Macro

NARA Format ID: NF00312

Extension(s):
- docm

Documentation:
- https://fileinfo.com/extension/docm
- https://docs.microsoft.com/en-us/deployoffice/compat/office-file-format-reference
- https://msdn.microsoft.com/en-us/library/gg134032.aspx
- https://www.lifewire.com/docm-file-2620747
- https://www.ncsc.gov.uk/guidance/macro-security-microsoft-office

Risk and Prioritization Analysis

☐ **Low Risk**

✓ Moderate Risk

☐ **High Risk**

14 Numeric Risk Rating

12 Numeric Prioritization Rating

Proposed Preservation Plan

✓ **Retain** file format in its existing format.

☐ **Transform** file to a new format.

☐ **Procure/develop tools** to preserve, manage and provide access to records of this type in their existing form.

☐ **Procure/develop tools** to transform the format to the preferred normalized form.

☐ **Provide Additional Information** so that the record type remains understandable/usable over time.

☐ **Explore Additional Options**

Justification: This file type is saved using the open XML-based Word document

Preferred Processing and Transformation Tool(s)

- Universal Office Converter - Command line library that converts between any document format supported by LibreOffice/OpenOffice (https://github.com/dagwieers/unoconv)
- Cloud Convert - https://cloudconvert.com/docm-to-pdf

Preferred Viewer/Access Software

- Adobe Acrobat

Open Document Format Document

NARA Format ID: NF00345

Extension(s):
- odt
- fodt

Documentation:
- https://en.wikipedia.org/wiki/OpenDocument
- http://docs.oasis-open.org/office/v1.1/OS/OpenDocument-v1.1.pdf
- http://docs.oasis-open.org/office/v1.2/OpenDocument-v1.2.pdf
- https://en.wikipedia.org/wiki/OpenDocument_technical_specification

Risk and Prioritization Analysis

☐ **Low Risk**

✓ Moderate Risk

☐ **High Risk**

14 Numeric Risk Rating

14 Numeric Prioritization Rating

Proposed Preservation Plan

✓ **Retain** file format in its existing format.

☐ **Transform** file to a new format.

☐ **Procure/develop tools** to preserve, manage and provide access to records of this type in their existing form.

☐ **Procure/develop tools** to transform the format to the preferred normalized form.

☐ **Provide Additional Information** so that the record type remains understandable/usable over time.

☐ **Explore Additional Options**

Justification: ODF is one of NARA's preferred formats.

Preferred Processing and Transformation Tool(s)

- Universal Office Converter - Command line library that converts between any document format supported by LibreOffice/OpenOffice (https://github.com/dagwieers/unoconv)

Preferred Viewer/Access Software

- ODF is a preferred format

Open Document Format Formula

NARA Format ID: NF00346

- odf

Documentation:
- https://en.wikipedia.org/wiki/OpenDocument
- http://docs.oasis-open.org/office/v1.1/OS/OpenDocument-v1.1.pdf
- http://docs.oasis-open.org/office/v1.2/OpenDocument-v1.2.pdf
- https://en.wikipedia.org/wiki/OpenDocument_technical_specification

Risk and Prioritization Analysis

✓ **Low Risk**
☐ **Moderate Risk**
☐ **High Risk**
45 Numeric Risk Rating
45 Numeric Prioritization Rating

Proposed Preservation Plan

- ✓ **Retain** file format in its existing format.
- ☐ **Transform** file to a new format.
- ☐ **Procure/develop tools** to preserve, manage and provide access to records of this type in their existing form.
- ☐ **Procure/develop tools** to transform the format to the preferred normalized form.
- ☐ **Provide Additional Information** so that the record type remains understandable/usable over time.
- ☐ **Explore Additional Options**

Justification: ODF is one of NARA's preferred formats.

Preferred Processing and Transformation Tool(s)

- Universal Office Converter - Command line library that converts between any document format supported by LibreOffice/OpenOffice (https://github.com/dagwieers/unoconv)

Preferred Viewer/Access Software

- ODF is a preferred format

Microsoft Word for Windows 2007

NARA Format ID: NF00311

Extension(s):
- doc
- docx

Documentation:
- https://msdn.microsoft.com/en-us/library/ee908652%28v=office.12%29
- https://en.wikipedia.org/wiki/History_of_Microsoft_Word#Word_2007
- https://en.wikipedia.org/wiki/Shared_Source_Initiative
- https://en.wikipedia.org/wiki/Microsoft_Office_2007
- https://www.loc.gov/preservation/digital/formats/intro/specifications.shtml

Risk and Prioritization Analysis

✓ **Low Risk**

☐ **Moderate Risk**
☐ **High Risk**
20 Numeric Risk Rating
18 Numeric Prioritization Rating

Proposed Preservation Plan

✓ **Retain** file format in its existing format.

✓ **Transform** file to a new format.

☐ **Procure/develop tools** to preserve, manage and provide access to records of this type in their existing form.
☐ **Procure/develop tools** to transform the format to the preferred normalized form.
☐ **Provide Additional Information** so that the record type remains understandable/usable over time.
☐ **Explore Additional Options**

Justification: .docx is one of NARA's acceptable formats. Transform .doc files to .docx.

Preferred Processing and Transformation Tool(s)

- Current supported version Microsoft Word

Preferred Viewer/Access Software

- Current supported version Microsoft Word

Microsoft Word for Windows 2.0

NARA Format ID: NF00310

Extension(s):
- doc

Documentation:
- https://en.wikipedia.org/wiki/Windows_2.0
- https://msdn.microsoft.com/en-us/library/office/cc313153%28v=office.12%29.aspx
- https://en.wikipedia.org/wiki/History_of_Microsoft_Word#Word_2007
- https://en.wikipedia.org/wiki/Windows_2.0

Risk and Prioritization Analysis

☐ **Low Risk**

✓ **Moderate Risk**

☐ **High Risk**
-3 Numeric Risk Rating
-5 Numeric Prioritization Rating

Proposed Preservation Plan

☐ **Retain** file format in its existing format.

✓ **Transform** file to a new format.
Transform into PDF

☐ **Procure/develop tools** to preserve, manage and provide access to records of this type in their existing form.

✓ **Procure/develop tools** to transform the format to the preferred normalized form.

☐ **Provide Additional Information** so that the record type remains understandable/usable over time.

☐ **Explore Additional Options**

Justification: This iteration of MS Word is not longer supported. Increasingly, migration to modern iterations of MS Word is difficult and damages the integrity of the original document.

Preferred Processing and Transformation Tool(s)

- Universal Office Converter - Command line library that converts between any document format supported by LibreOffice/OpenOffice (https://github.com/dagwieers/unoconv)

Preferred Viewer/Access Software

- Adobe Acrobat for PDF

Microsoft Word Backup File, versions 97 onwards

NARA Format ID: NF00304

Extension(s):
- wbk

Documentation:
- https://fileinfo.com/extension/wbk
- http://fileformats.archiveteam.org/wiki/Microsoft_Word

Risk and Prioritization Analysis

☐ **Low Risk**

✓ **Moderate Risk**

☐ **High Risk**

9 Numeric Risk Rating

7 Numeric Prioritization Rating

Proposed Preservation Plan

✓ **Retain** file format in its existing format.

✓ **Transform** file to a new format.

   Restore to Word files as appropriate

☐ **Procure/develop tools** to preserve, manage and provide access to records of this type in their existing form.

☐ **Procure/develop tools** to transform the format to the preferred normalized form.

☐ **Provide Additional Information** so that the record type remains understandable/usable over time.

✓ **Explore Additional Options**

Justification: Supported by current Microsoft Word versions. Determine whether the Word Backup files are associated with Word files and if the files should be considered record or non-record before taking any potential preservation actions, including restoring as Word .docx files.

Preferred Processing and Transformation Tool(s)

- Universal Office Converter - Command line library that converts between any document format supported by LibreOffice/OpenOffice (https://github.com/dagwieers/unoconv)

Preferred Viewer/Access Software

- Adobe Acrobat for PDF

EndNote Import File

NARA Format ID: NF00172

Extension(s):
- enw

Documentation:
- https://en.wikipedia.org/wiki/EndNote

Risk and Prioritization Analysis

☐ **Low Risk**

✓ **Moderate Risk**

☐ **High Risk**

14 Numeric Risk Rating

14 Numeric Prioritization Rating

Proposed Preservation Plan

☐ **Retain** file format in its existing format.

✓ **Transform** file to a new format.

      TBD, preferably ASCII delimited text

☐ **Procure/develop tools** to preserve, manage and provide access to records of this type in their existing form.

✓ **Procure/develop tools** to transform the format to the preferred normalized form.

☐ **Provide Additional Information** so that the record type remains understandable/usable over time.

☐ **Explore Additional Options**

Justification: Import files are already structured plain text that follow the EndNote schema with a proprietary extension, so conversion will not alter the content.

Preferred Processing and Transformation Tool(s)

- TBD

Preferred Viewer/Access Software

- Any Text Editor or Word Processing program.

OmniPage Pro OCR

NARA Format ID: NF00344

Extension(s):
- opd

Documentation:
- https://en.wikipedia.org/wiki/OmniPage
- http://fileformats.archiveteam.org/wiki/Ext:met

Risk and Prioritization Analysis

☐ **Low Risk**

✓ **Moderate Risk**

☐ **High Risk**
17 Numeric Risk Rating
14 Numeric Prioritization Rating

Proposed Preservation Plan

Recommended actions (more than one can be selected):

☐ **Retain** file format in its existing format.

✓ **Transform** file to a new format.
> Transform into ASCII text

☐ **Procure/develop tools** to preserve, manage and provide access to records of this type in their existing form.

✓ **Procure/develop tools** to transform the format to the preferred normalized form.

☐ **Provide Additional Information** so that the record type remains understandable/usable over time.

☐ **Explore Additional Options**

Justification: OCR files are already structured plain text that follow the OmniPage structure with a proprietary extension, so conversion will not alter the content.

Preferred Processing and Transformation Tool(s)

- TBD

Preferred Viewer/Access Software

- Any Text Editor or Word Processing program.

eXtensible Markup Language (xml), Document Type Definition (DTD), eXtensible Markup Language Schema (xsd), and eXtensible Style Language (xsl, xslt)

Marked-up plain text Unicode files which can be used to represent, parse, format, and display structured text.

SEE: Web Content Record Format Preservation Action Plan