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

Plan Date: 20231122
Template: 202105

Textual and Word Processing

Textual and Word Processing records are 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.

Significant Properties of Textual and Word Processing

The variety of agency usage and types of textual and word processing formats limit NARA's ability to fully define consistent properties, especially appearance. Built-in tools such as macros or externally linked files are related to other record categories which have their own significant properties.

Some appearance properties are discussed in the Structure Properties section, since the appearance of a textual record can be partially based on the application used to create the record. 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 may not be as significant.

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 properties 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?

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 property. 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 properties 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 properties 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	Many textual and word processing file formats are generated using multiple components that are	

	specified via a set of structured data relationships, which may include: • 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.	Significant properties are currently dependent on the export format. If macros or other linkages are present, evaluate for functionality before considering removing the features or transforming the file.

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 Textual and Word Processing
[Bulletin 2014-04](#)

- Preferred:
  - ASCII Text (7-bit)
  - Unicode Text (UTF-8, UTF-16)
  - OpenDocument Text Format (ODF, all versions)
  - PDF/A-1
  - PDF/A-2

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

Current NARA Format(s) for Public Access and Reference for Textual and Word Processing

Formats for Public Access are those made available online through the National Archives Catalog. Formats for Reference are defined as those made available to researchers upon direct requests for digital copies.

Formats Available for Public Access: PDF, DOCX, ASCII Text, ODT

Format(s) Available for Reference: Where possible, records are delivered to researchers in the formats in which they are preserved.