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

Plan Date: 20210525
Template: 202105

## Textual and Word Processing Records

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.

## Significant Properties of Textual/Word Processing Records

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

Some appearance properties are discussed in the Structure Properties 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 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?

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 significant properties.

## 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.<br>● Typeface (Arial, Times New Roman, etc.)<br>● Size (10 pt, 18 pt, etc.)<br>● Pitch<br>● Spacing<br>● 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.<br>● Hue: color family or name<br>● Saturation: purity or | Color is essential if it bears meaning and/or value, or if the appearance is fixed. |

| | | |
|---|---|---|
| | sharpness<br>● Brightness: shade or tint<br>● 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<br>● Arrays<br>● Models<br>● Trees<br>● Tables | |

|  |  |  |
|---|---|---|
|  | - Linkage: connection between or within records. If connections exist, then they are core<br>- 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. |

## 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/Word Processing Records**
Bulletin 2014-04

- 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 Format(s) for Public Access and Reference for Textual/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: Content created or delivered for public access in the Catalog is delivered primarily in the following file formats: PDF (Textual and Image), JPEG (Textual and Image), MP3 (Audio), and MP4 (Audio/Video) and ASCII (Datasets). Other file formats may be present depending on when they were added to the Catalog.

Format(s) Available for Reference: PDF