

Preservation Action Plan: Presentation and Publishing Formats National Archives and Records Administration (NARA)

Plan Date: 20200629

Template: 201907

Electronic Record or Digital Surrogate Types and Associated Formats

Publishing formats include a variety of files related to desktop publishing software used to create graphically rich documents, magazines, books, brochures, newspapers and other more sophisticated texts. Besides file formats of the publications themselves, this record type includes file formats associated with rendering files, such as printer control language and page description language files.

Essential Characteristics of Presentation and Publishing Formats

The rapid evolution of textual record types into multimedia formats, 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. 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.

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?

Appearance

| Name | Definition | Function Description |
|--------------------|--|--|
| Language | A language identifier specifying the natural language(s) used in the document. | |
| Fonts | <p>A list of fonts used in the document, and an indication of whether or not a font is embedded in a document.</p> <ul style="list-style-type: none"> 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 | <p>Identification of the use of color in text and layout elements, e.g. borders, boxes.</p> <ul style="list-style-type: none"> Hue: color family or name Saturation: purity or | Color is essential if it bears meaning and/or value, or if the appearance is fixed. |

| | | |
|--------------|---|--|
| | sharpness <ul style="list-style-type: none"> • 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. | A table is an arrangement of <u>data</u> in rows and columns, or possibly in a more complex structure. (Wikipedia) |
| Graphics | Images embedded in the document, or graphics created using a word processing graphics feature. | Used to point readers and viewers to particular information. They are also used to supplement text in an effort to aid readers in their understanding of a particular concept or make the concept more clear or interesting. (Wikipedia) |
| Outline | Includes content in an outline structure. | An outline, also called a hierarchical outline, is a list arranged to show hierarchical relationships and is a type of tree structure. An outline is used to present the main points or topics of a given subject. Each item in an outline may be divided into additional sub-items. (Wikipedia) |
| 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 | <ul style="list-style-type: none">• 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 | | 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 Presentation and Publishing Formats

[Bulletin 2014-04](#)

- Preferred:
 - OpenDocument Text Format (ODF) 1.0
 - Portable Document Format/Archival (PDF/A)
 - Tagged Image File Format (TIFF)
 - JPEG2000 (JP2)
 - Portable Network Graphics (PNG)
- Acceptable:
 - PDF 1.0-1.7
 - JPEG File Interchange Format (JFIF) with Joint Photographic Experts Group (JPEG) compression
 - Graphics Interchange Format (GIF)
 - PDF/A-2

Current NARA Format(s) for Public Access and Reference for Presentation and Publishing Formats

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, TIFF, JP2