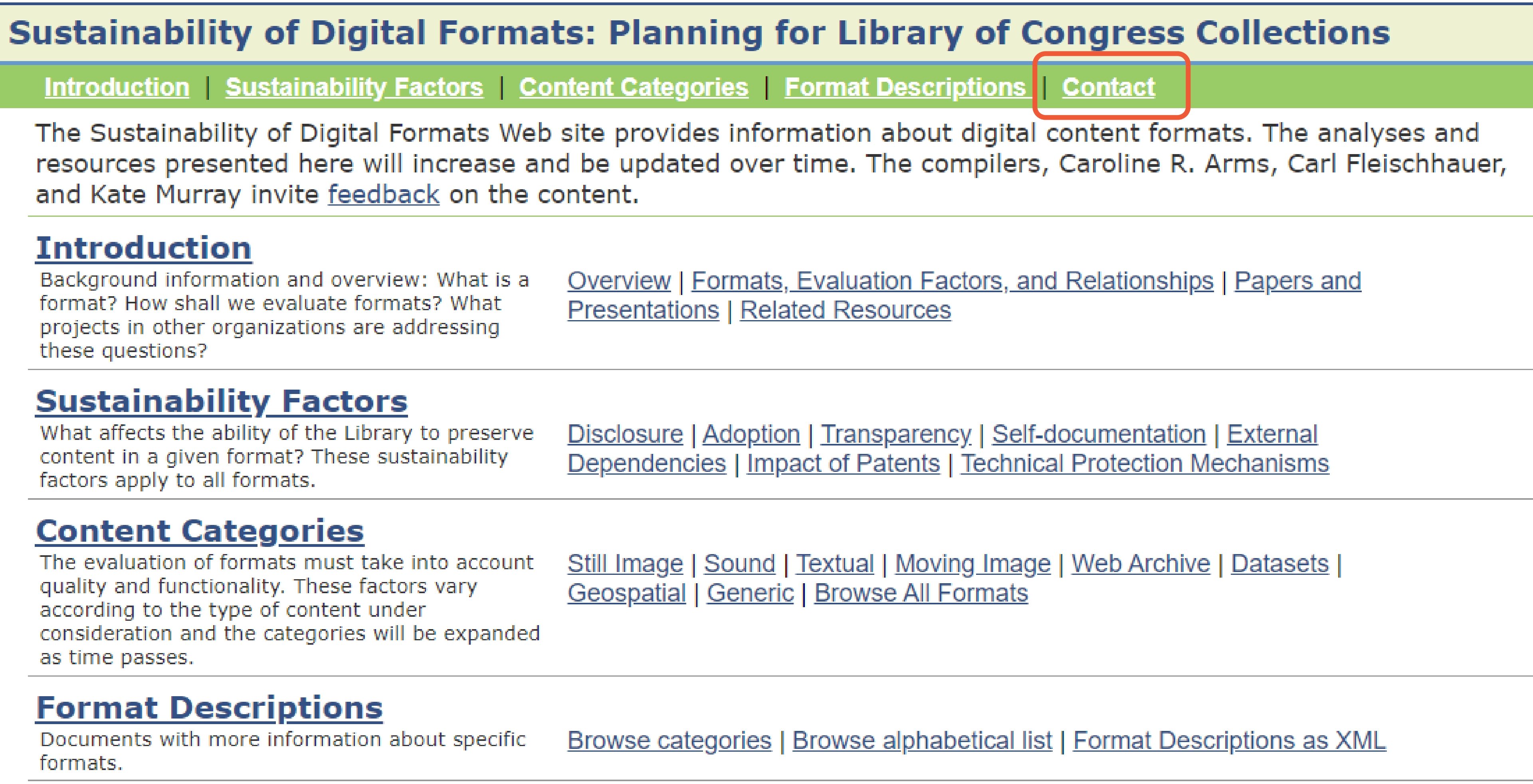
LIBRARY LIBRARY OF CONGRESS

February 9, 2021

Digital File Format Resources at the Library of Congress

DPC Briefing Day: Preservation Planning and Technology Watch

Kate Murray
kmur@loc.gov



Sustainability of Digital Formats: Planning for Library of Congress Collections

Introduction | Sustainability Factors | Content Categories | Format Descriptions | Contact

The Sustainability of Digital Formats Web site provides information about digital content formats. The analyses and resources presented here will increase and be updated over time. The compilers, Caroline R. Arms, Carl Fleischhauer, and Kate Murray invite feedback on the content.

Introduction

Background information and overview: What is a format? How shall we evaluate formats? What projects in other organizations are addressing these questions?

Overview | Formats, Evaluation Factors, and Relationships | Papers and Presentations | Related Resources

Sustainability Factors

What affects the ability of the Library to preserve content in a given format? These sustainability factors apply to all formats.

Disclosure | Adoption | Transparency | Self-documentation | External Dependencies | Impact of Patents | Technical Protection Mechanisms

Content Categories

The evaluation of formats must take into account quality and functionality. These factors vary according to the type of content under consideration and the categories will be expanded as time passes.

Still Image | Sound | Textual | Moving Image | Web Archive | Datasets | Geospatial | Generic | Browse All Formats

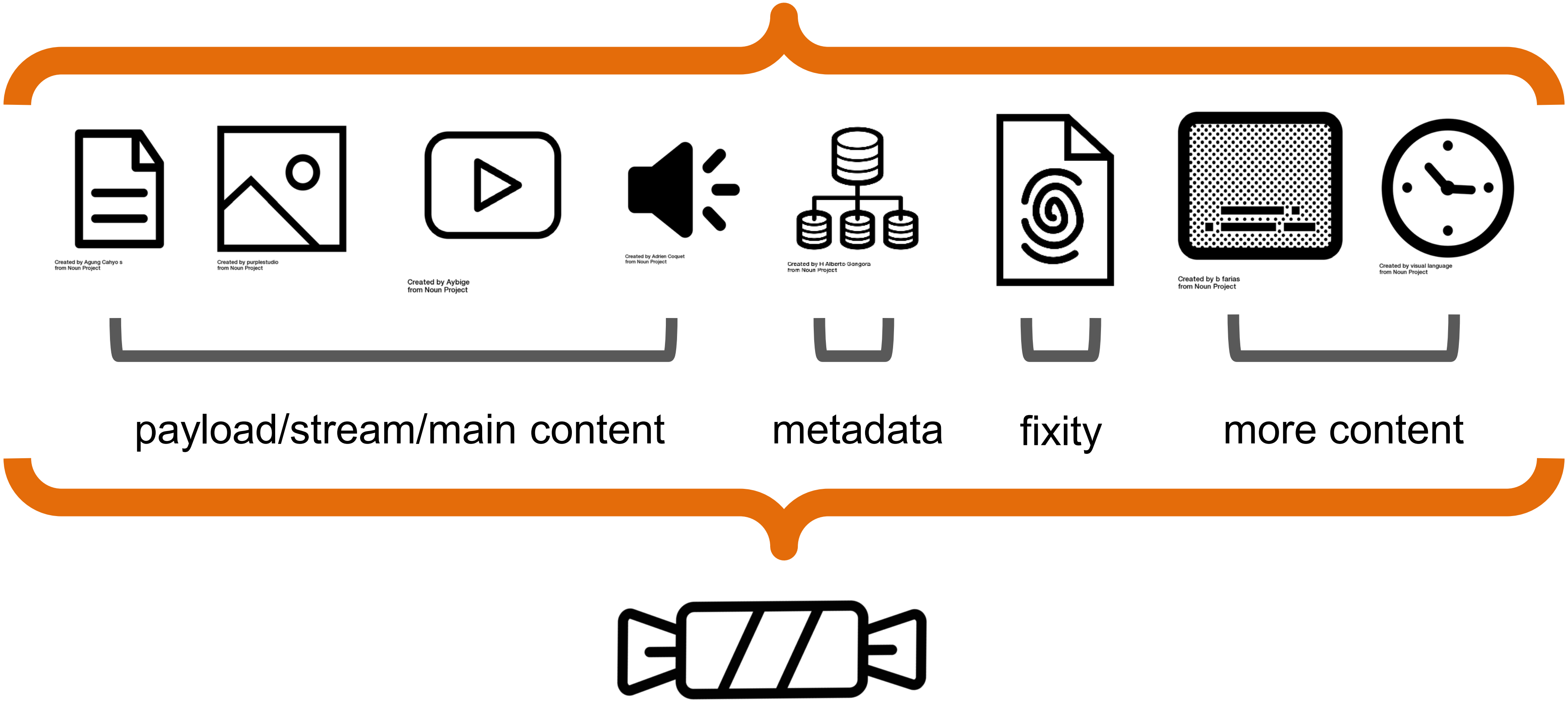
Format Descriptions

Documents with more information about specific formats.

Browse categories | Browse alphabetical list | Format Descriptions as XML

http://www.loc.gov/preservation/digital/formats/index.html

what’s in a file?



payload/stream/main content

metadata

fixity

more content

Created by Agung Cahyo s
from Noun Project

Created by purplestudio
from Noun Project

Created by Aybige
from Noun Project

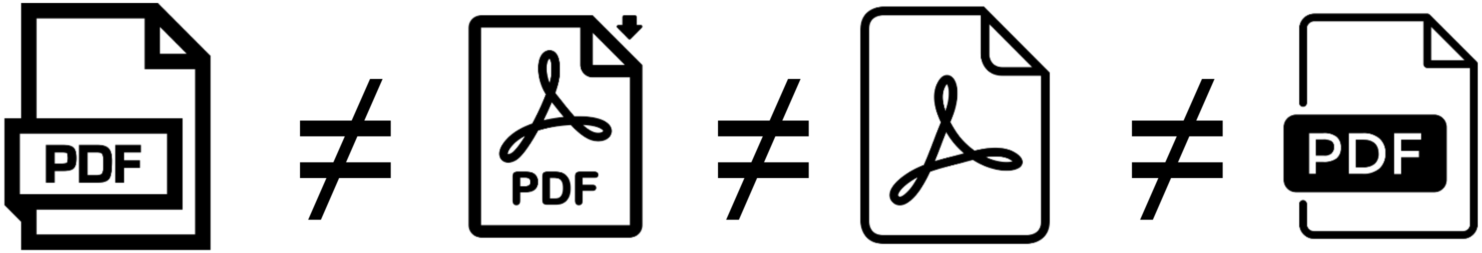
Created by Adrien Coquet
from Noun Project

Created by H Alberto Gongora
from Noun Project

Created by b farias
from Noun Project

Created by visual language
from Noun Project

Created by Sneha
from Noun Project



Created by Alfredo @ IconsAlfredo.com
from Noun Project

Created by Yamini Ahluwalia
from Noun Project

Created by Rediffusion
from Noun Project

Created by Ilham Fitrotul Hayat
from Noun Project

even files that look the same, with the same file extension, can be very different

pdf

mother of all container formats

its flexibility is its curse

pdf/a (“archival” friendly. maybe? kinda?)

no encryption

no javascript

audio/video only permitted in pdf/a-3

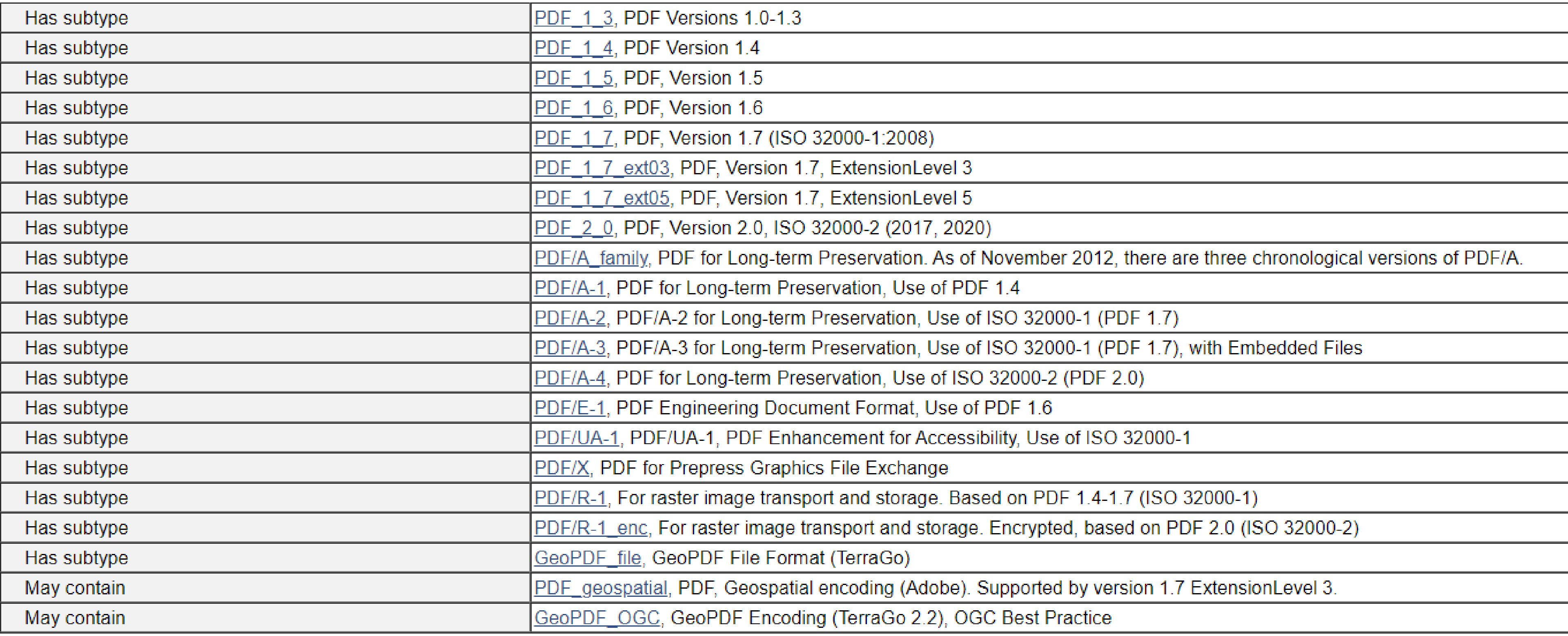
pdf/x

pdf/ua

relationships are hard.

has subtype | subtype of | contains | may contain | used by | must have component | may have component | component of | defined via | requires | modification of | has modified version | extension of | has extension | has earlier version | has later version | version of | equivalent to | affinity to | additional | other | yada yada yada ...

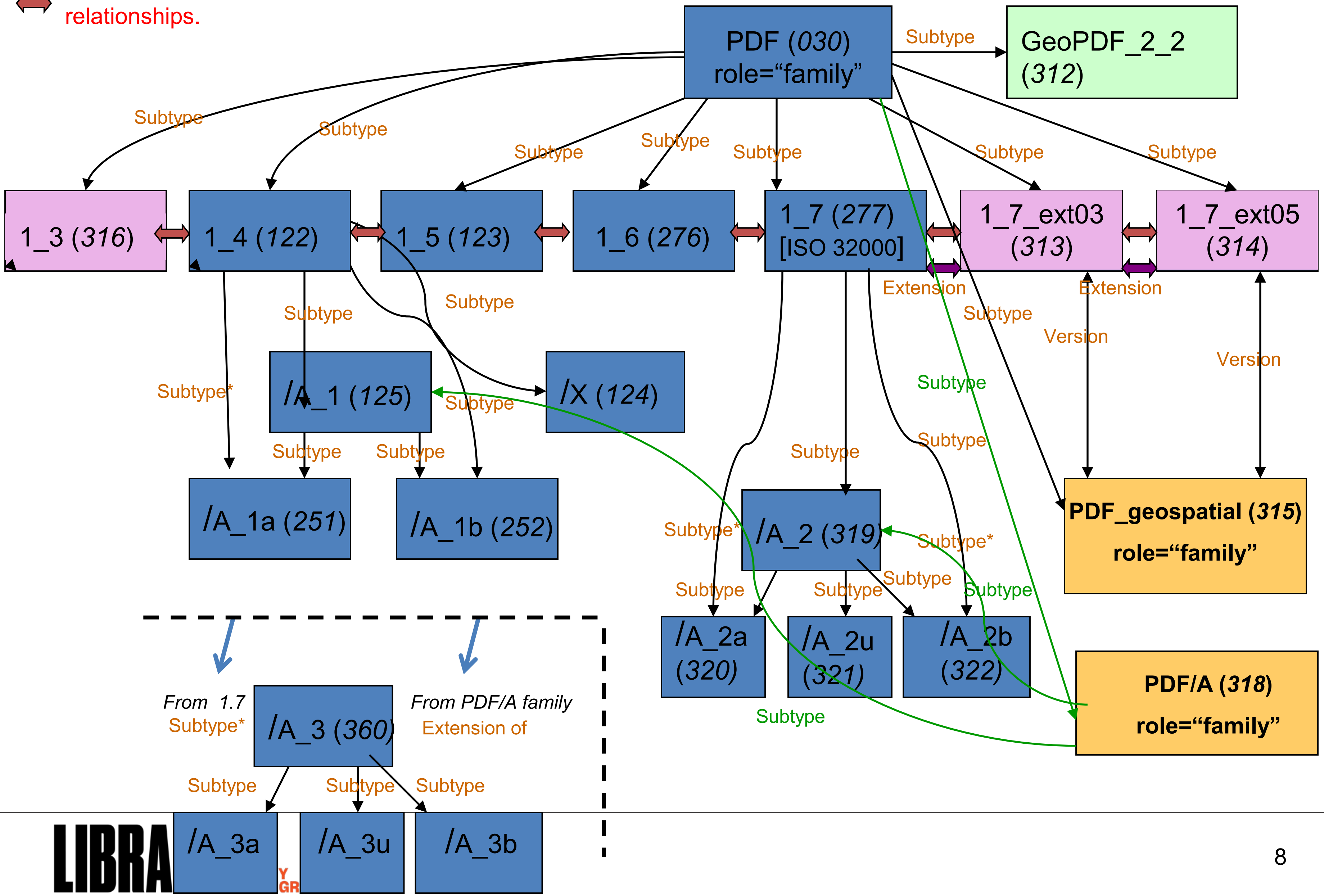
PDF is much more than text -- a file format, a wrapper, a bundling format, all in one – note complexity of relationships



Has subtype	PDF_1_3, PDF Versions 1.0-1.3
Has subtype	PDF_1_4, PDF Version 1.4
Has subtype	PDF_1_5, PDF, Version 1.5
Has subtype	PDF_1_6, PDF, Version 1.6
Has subtype	PDF_1_7, PDF, Version 1.7 (ISO 32000-1:2008)
Has subtype	PDF_1_7_ext03, PDF, Version 1.7, ExtensionLevel 3
Has subtype	PDF_1_7_ext05, PDF, Version 1.7, ExtensionLevel 5
Has subtype	PDF_2_0, PDF, Version 2.0, ISO 32000-2 (2017, 2020)
Has subtype	PDF/A_family, PDF for Long-term Preservation. As of November 2012, there are three chronological versions of PDF/A.
Has subtype	PDF/A-1, PDF for Long-term Preservation, Use of PDF 1.4
Has subtype	PDF/A-2, PDF/A-2 for Long-term Preservation, Use of ISO 32000-1 (PDF 1.7)
Has subtype	PDF/A-3, PDF/A-3 for Long-term Preservation, Use of ISO 32000-1 (PDF 1.7), with Embedded Files
Has subtype	PDF/A-4, PDF for Long-term Preservation, Use of ISO 32000-2 (PDF 2.0)
Has subtype	PDF/E-1, PDF Engineering Document Format, Use of PDF 1.6
Has subtype	PDF/UA-1, PDF/UA-1, PDF Enhancement for Accessibility, Use of ISO 32000-1
Has subtype	PDF/X, PDF for Prepress Graphics File Exchange
Has subtype	PDF/R-1, For raster image transport and storage. Based on PDF 1.4-1.7 (ISO 32000-1)
Has subtype	PDF/R-1_enc, For raster image transport and storage. Encrypted, based on PDF 2.0 (ISO 32000-2)
Has subtype	GeoPDF_file, GeoPDF File Format (TerraGo)
May contain	PDF_geospatial, PDF, Geospatial encoding (Adobe). Supported by version 1.7 ExtensionLevel 3.
May contain	GeoPDF_OGC, GeoPDF Encoding (TerraGo 2.2), OGC Best Practice

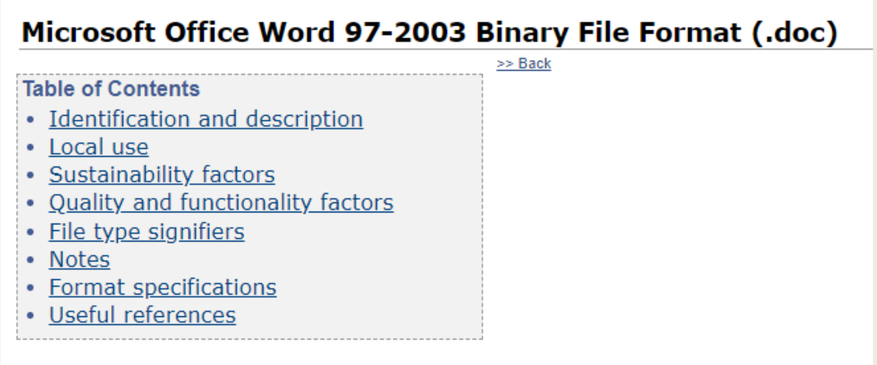
Adobe PDF lineage, with FDD number in italics.

⟺ Indicates earlier/later version relationships.



FDD

format description document



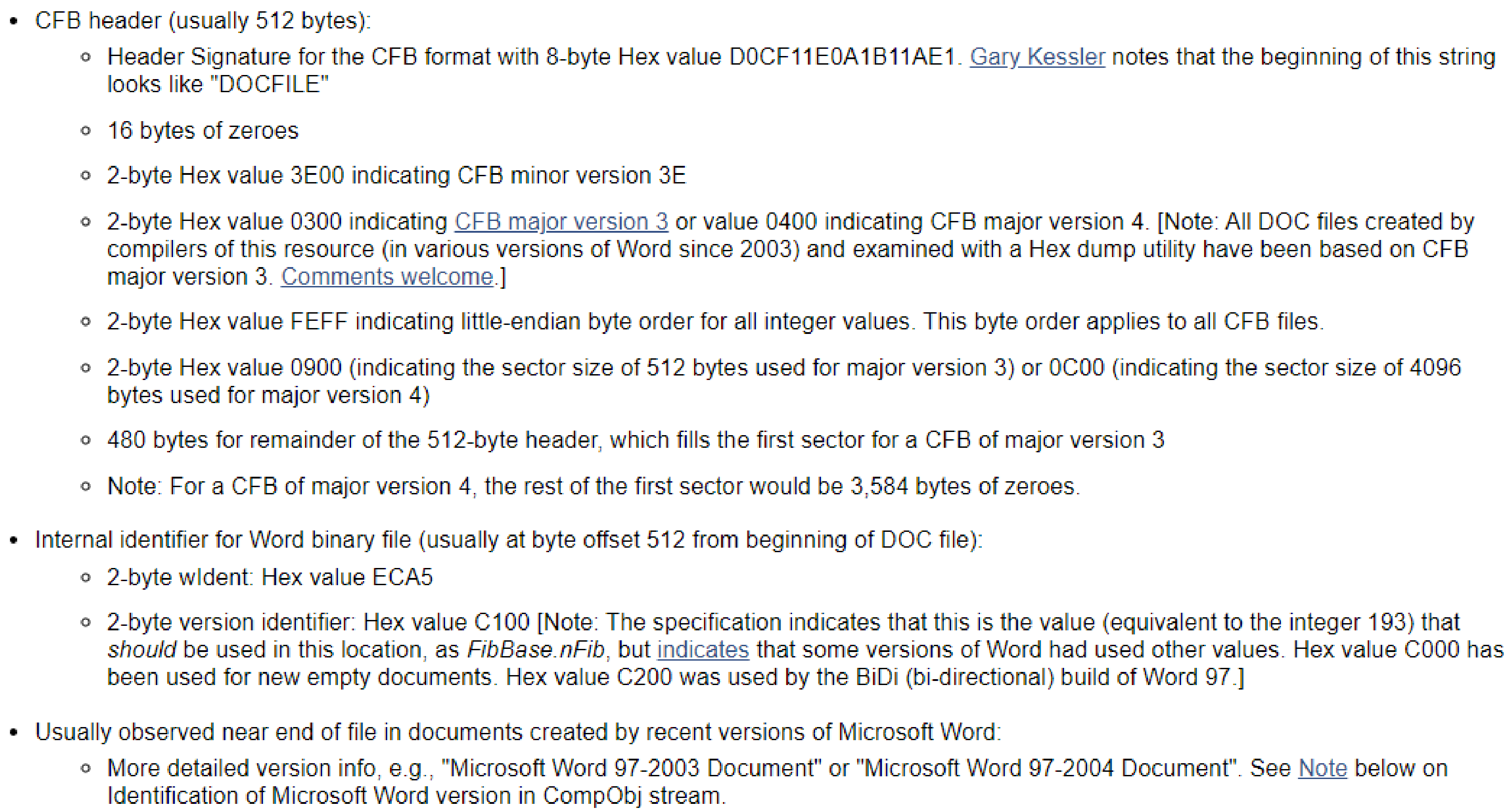
Microsoft Office Word 97-2003 Binary File Format (.doc)

>> Back

Table of Contents

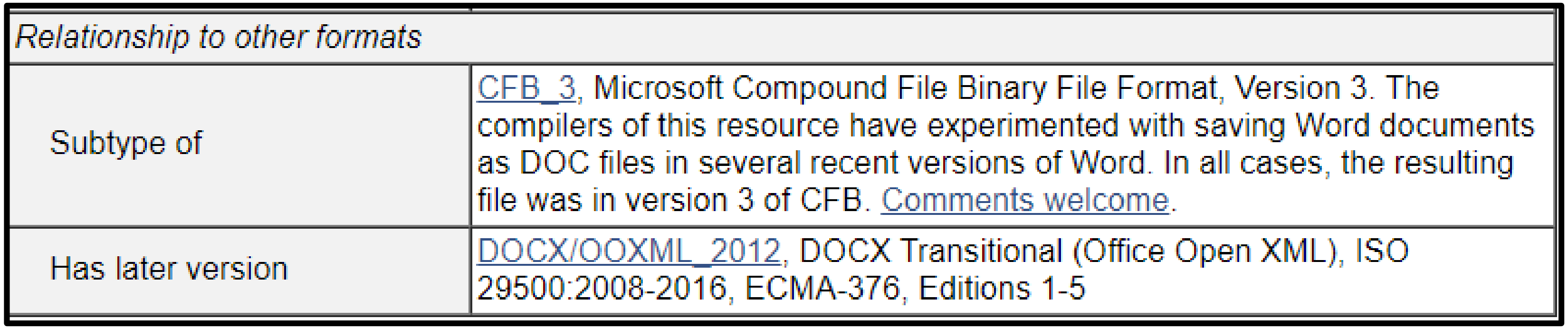
- Identification and description
- Local use
- Sustainability factors
- Quality and functionality factors
- File type signifiers
- Notes
- Format specifications
- Useful references

http://www.loc.gov/preservation/digital/formats/fdd/fdd000509.shtml



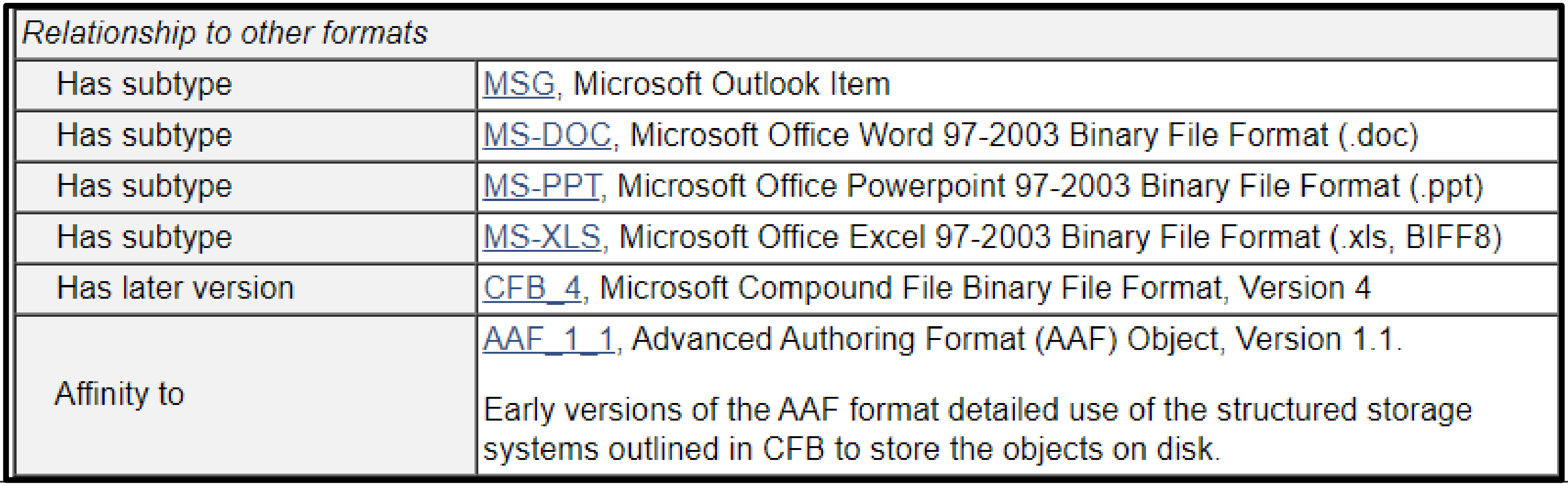
- CFB header (usually 512 bytes):
  - Header Signature for the CFB format with 8-byte Hex value D0CF11E0A1B11AE1. Gary Kessler notes that the beginning of this string looks like "DOCFILE"
  - 16 bytes of zeroes
  - 2-byte Hex value 3E00 indicating CFB minor version 3E
  - 2-byte Hex value 0300 indicating CFB major version 3 or value 0400 indicating CFB major version 4. [Note: All DOC files created by compilers of this resource (in various versions of Word since 2003) and examined with a Hex dump utility have been based on CFB major version 3. Comments welcome.]
  - 2-byte Hex value FEFF indicating little-endian byte order for all integer values. This byte order applies to all CFB files.
  - 2-byte Hex value 0900 (indicating the sector size of 512 bytes used for major version 3) or 0C00 (indicating the sector size of 4096 bytes used for major version 4)
  - 480 bytes for remainder of the 512-byte header, which fills the first sector for a CFB of major version 3
  - Note: For a CFB of major version 4, the rest of the first sector would be 3,584 bytes of zeroes.
- Internal identifier for Word binary file (usually at byte offset 512 from beginning of DOC file):
  - 2-byte wIdent: Hex value ECA5
  - 2-byte version identifier: Hex value C100 [Note: The specification indicates that this is the value (equivalent to the integer 193) that *should* be used in this location, as *FibBase.nFib*, but indicates that some versions of Word had used other values. Hex value C000 has been used for new empty documents. Hex value C200 was used by the BiDi (bi-directional) build of Word 97.]
- Usually observed near end of file in documents created by recent versions of Microsoft Word:
  - More detailed version info, e.g., "Microsoft Word 97-2003 Document" or "Microsoft Word 97-2004 Document". See Note below on Identification of Microsoft Word version in CompObj stream.

description



Relationship to other formats	
Subtype of	CFB_3, Microsoft Compound File Binary File Format, Version 3. The compilers of this resource have experimented with saving Word documents as DOC files in several recent versions of Word. In all cases, the resulting file was in version 3 of CFB. Comments welcome.
Has later version	DOCX/OOXML_2012, DOCX Transitional (Office Open XML), ISO 29500:2008-2016, ECMA-376, Editions 1-5

MS-DOC (509) is based on CFB 3 (380) but CFB 3 is the basis for MANY other formats – all the old Microsoft formats – not just MS-DOC



Relationship to other formats	
Has subtype	MSG, Microsoft Outlook Item
Has subtype	MS-DOC, Microsoft Office Word 97-2003 Binary File Format (.doc)
Has subtype	MS-PPT, Microsoft Office Powerpoint 97-2003 Binary File Format (.ppt)
Has subtype	MS-XLS, Microsoft Office Excel 97-2003 Binary File Format (.xls, BIFF8)
Has later version	CFB_4, Microsoft Compound File Binary File Format, Version 4
Affinity to	AAF_1_1, Advanced Authoring Format (AAF) Object, Version 1.1. Early versions of the AAF format detailed use of the structured storage systems outlined in CFB to store the objects on disk.

2 types of evaluation factors

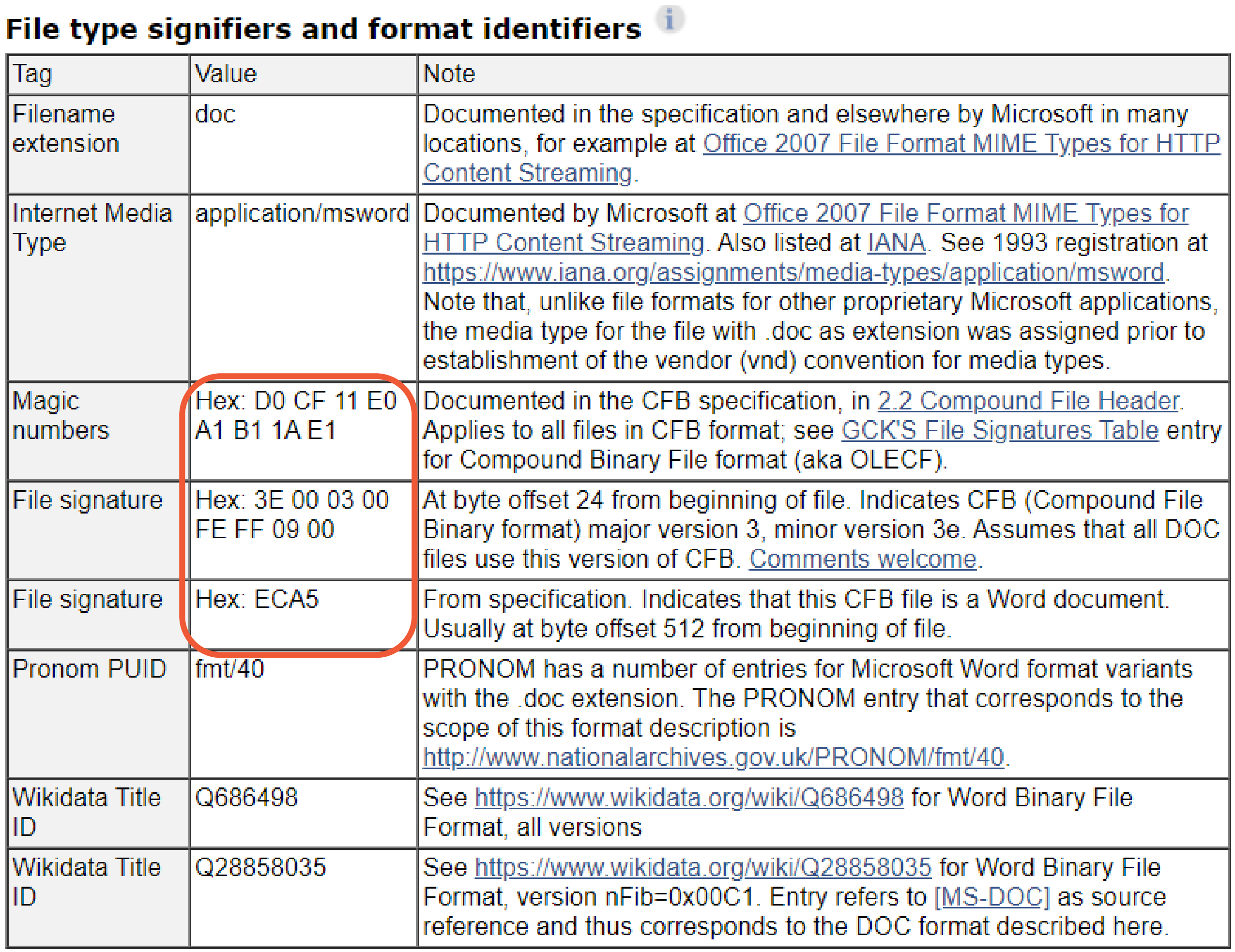
sustainability factors for all formats

influence feasibility and cost of preserving content in the face of future change

quality & functionality factors that vary by content category

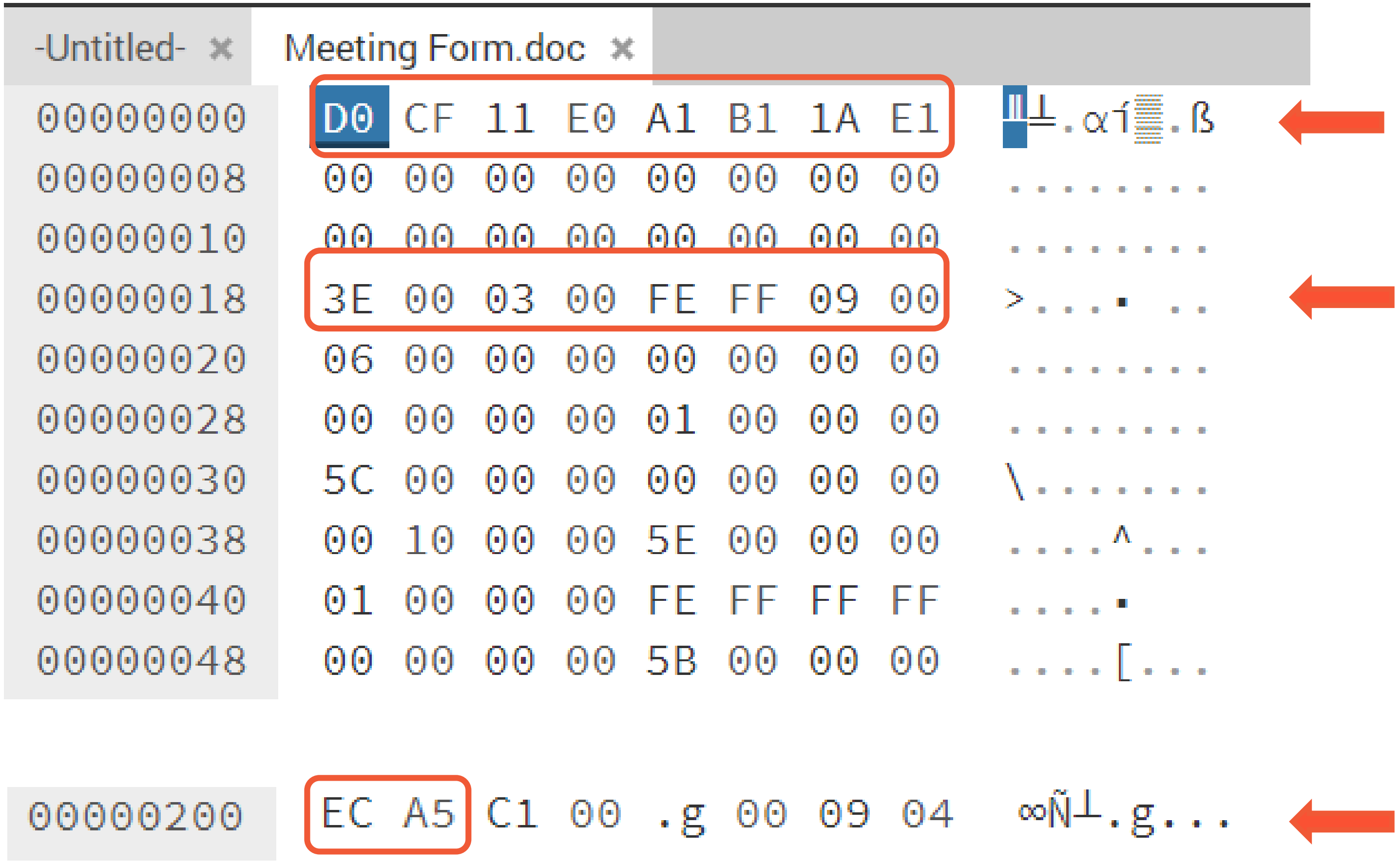
reflect considerations that will be expected by future users

kate’s favorite section: identifiers



File type signifiers and format identifiers

Tag	Value	Note
Filename extension	doc	Documented in the specification and elsewhere by Microsoft in many locations, for example at Office 2007 File Format MIME Types for HTTP Content Streaming.
Internet Media Type	application/msword	Documented by Microsoft at Office 2007 File Format MIME Types for HTTP Content Streaming. Also listed at IANA. See 1993 registration at https://www.iana.org/assignments/media-types/application/msword. Note that, unlike file formats for other proprietary Microsoft applications, the media type for the file with .doc as extension was assigned prior to establishment of the vendor (vnd) convention for media types.
Magic numbers	Hex: D0 CF 11 E0 A1 B1 1A E1	Documented in the CFB specification, in 2.2 Compound File Header. Applies to all files in CFB format; see GCK'S File Signatures Table entry for Compound Binary File format (aka OLECF).
File signature	Hex: 3E 00 03 00 FE FF 09 00	At byte offset 24 from beginning of file. Indicates CFB (Compound File Binary format) major version 3, minor version 3e. Assumes that all DOC files use this version of CFB. Comments welcome.
File signature	Hex: ECA5	From specification. Indicates that this CFB file is a Word document. Usually at byte offset 512 from beginning of file.
Pronom PUID	fmt/40	PRONOM has a number of entries for Microsoft Word format variants with the .doc extension. The PRONOM entry that corresponds to the scope of this format description is http://www.nationalarchives.gov.uk/PRONOM/fmt/40.
Wikidata Title ID	Q686498	See https://www.wikidata.org/wiki/Q686498 for Word Binary File Format, all versions
Wikidata Title ID	Q28858035	See https://www.wikidata.org/wiki/Q28858035 for Word Binary File Format, version nFib=0x00C1. Entry refers to [MS-DOC] as source reference and thus corresponds to the DOC format described here.

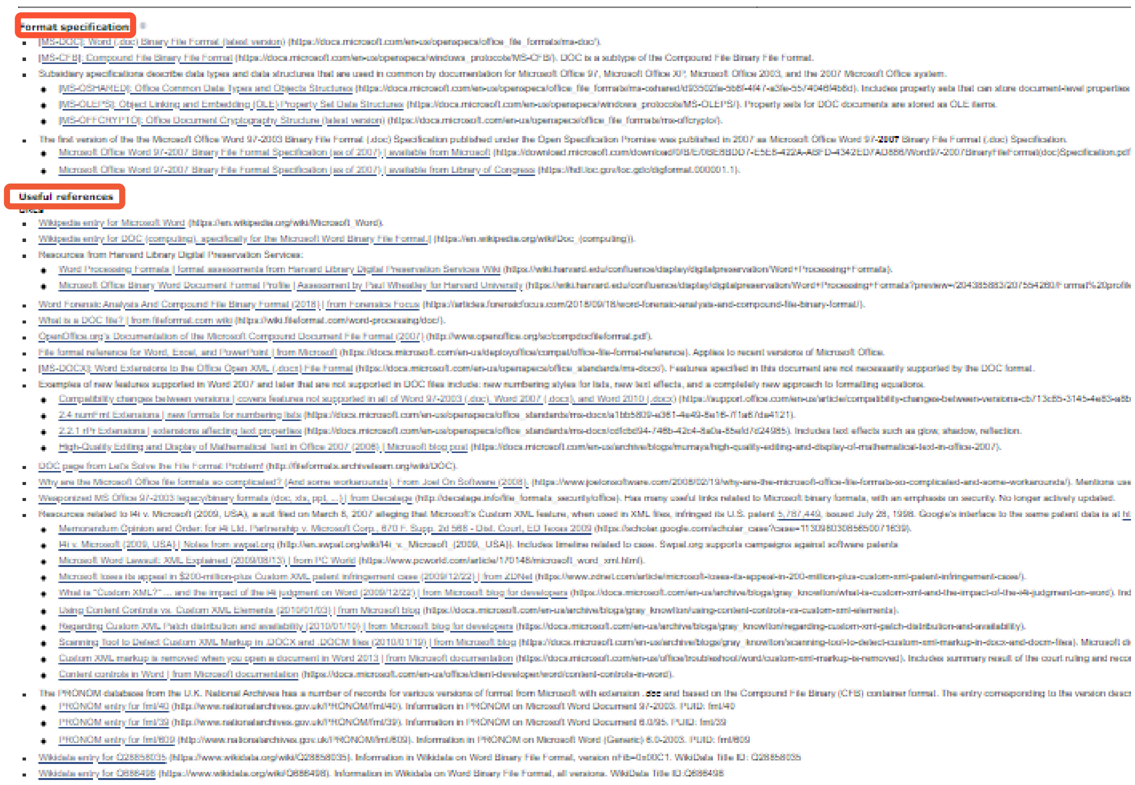


```
00000000  D0 CF 11 E0 A1 B1 1A E1   ╨╧.αí▒.ß
00000008  00 00 00 00 00 00 00 00   ........
00000010  00 00 00 00 00 00 00 00   ........
00000018  3E 00 03 00 FE FF 09 00   >...■  ..
00000020  06 00 00 00 00 00 00 00   ........
00000028  00 00 00 00 01 00 00 00   ........
00000030  5C 00 00 00 00 00 00 00   \.......
00000038  00 10 00 00 5E 00 00 00   ....^...
00000040  01 00 00 00 FE FF FF FF   ....■
00000048  00 00 00 00 5B 00 00 00   ....[...

00000200  EC A5 C1 00 .g 00 09 04   ∞Ñ┴.g...
```

hex view

lots and lots of references



*

Format specification

- [MS-DOC]: Word (.doc) Binary File Format (latest version) (https://docs.microsoft.com/en-us/openspecs/office_file_formats/ms-doc/).
- [MS-CFB]: Compound File Binary File Format (https://docs.microsoft.com/en-us/openspecs/windows_protocols/MS-CFB/). DOC is a subtype of the Compound File Binary File Format.
- Subsidiary specifications describe data types and data structures that are used in common by documentation for Microsoft Office 97, Microsoft Office XP, Microsoft Office 2003, and the 2007 Microsoft Office system.
  - [MS-OSHARED]: Office Common Data Types and Objects Structures (https://docs.microsoft.com/en-us/openspecs/office_file_formats/ms-oshared/d93502fe-566f-4f47-a3fe-55740484b6d). Includes property sets that can store document-level properties
  - [MS-OLEPS]: Object Linking and Embedding (OLE) Property Set Data Structures (https://docs.microsoft.com/en-us/openspecs/windows_protocols/MS-OLEPS/). Property sets for DOC documents are stored as OLE items.
  - [MS-OFFCRYPTO]: Office Document Cryptography Structure (latest version) (https://docs.microsoft.com/en-us/openspecs/office_file_formats/ms-offcrypto/).
- The first version of the the Microsoft Office Word 97-2003 Binary File Format (.doc) Specification published under the Open Specification Promise was published in 2007 as Microsoft Office Word 97-2007 Binary File Format (.doc) Specification.
  - Microsoft Office Word 97-2007 Binary File Format Specification (as of 2007) | available from Microsoft (https://download.microsoft.com/download/0/B/E/0BE8BDD7-E5E8-422A-ABFD-4342ED7AD886/Word97-2007BinaryFileFormat(doc)Specification.pdf
  - Microsoft Office Word 97-2007 Binary File Format Specification (as of 2007) | available from Library of Congress (https://hdl.loc.gov/loc.gdc/digformat.000001.1).

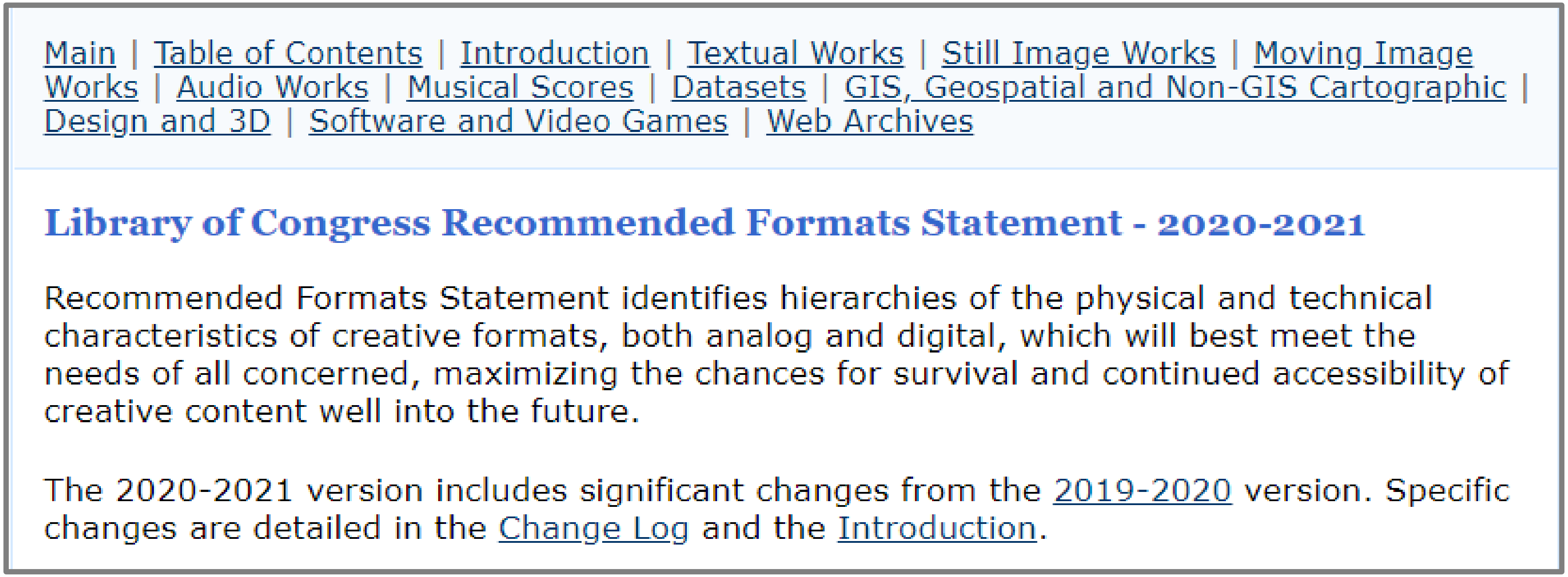
* *

Useful references

- Wikipedia entry for Microsoft Word (https://en.wikipedia.org/wiki/Microsoft_Word).
- Wikipedia entry for DOC (computing), specifically for the Microsoft Word Binary File Format.) (https://en.wikipedia.org/wiki/Doc_(computing)).
- Resources from Harvard Library Digital Preservation Services:
  - Word Processing Formats | format assessments from Harvard Library Digital Preservation Services Wiki (https://wiki.harvard.edu/confluence/display/digitalpreservation/Word+Processing+Formats).
  - Microsoft Office Binary Word Document Format Profile | Assessment by Paul Wheatley for Harvard University (https://wiki.harvard.edu/confluence/display/digitalpreservation/Word+Processing+Formats?preview=/204385883/207554260/Format%20profile
- Word Forensic Analysis And Compound File Binary Format (2018) | from Forensics Focus (https://articles.forensicfocus.com/2018/09/18/word-forensic-analysis-and-compound-file-binary-format/).
- What is a DOC file? | from fileformat.com wiki (https://wiki.fileformat.com/word-processing/doc/).
- OpenOffice.org's Documentation of the Microsoft Compound Document File Format (2007) (http://www.openoffice.org/sc/compdocfileformat.pdf).
- File format reference for Word, Excel, and PowerPoint | from Microsoft (https://docs.microsoft.com/en-us/deployoffice/compat/office-file-format-reference). Applies to recent versions of Microsoft Office.
- [MS-DOCX]: Word Extensions to the Office Open XML (.docx) File Format (https://docs.microsoft.com/en-us/openspecs/office_standards/ms-docx/). Features specified in this document are not necessarily supported by the DOC format.
- Examples of new features supported in Word 2007 and later that are not supported in DOC files include: new numbering styles for lists, new text effects, and a completely new approach to formatting equations.
  - Compatibility changes between versions | covers features not supported in all of Word 97-2003 (.doc), Word 2007 (.docx), and Word 2010 (.docx) (https://support.office.com/en-us/article/compatibility-changes-between-versions-cb713c65-3145-4e53-a8b
  - 2.4 numFmt Extensions | new formats for numbering lists (https://docs.microsoft.com/en-us/openspecs/office_standards/ms-docx/a1bb5809-e381-4e49-8a16-71a87da4121).
  - 2.2.1 rPr Extensions | extensions affecting text properties (https://docs.microsoft.com/en-us/openspecs/office_standards/ms-docx/cdfcbd94-748b-42c4-8a0a-85efd7d24985). Includes text effects such as glow, shadow, reflection.
  - High-Quality Editing and Display of Mathematical Text in Office 2007 (2006) | Microsoft blog post (https://docs.microsoft.com/en-us/archive/blogs/murrays/high-quality-editing-and-display-of-mathematical-text-in-office-2007).
- DOC page from Let's Solve the File Format Problem (http://fileformats.archiveteam.org/wiki/DOC).
- Why are the Microsoft Office file formats so complicated? (And some workarounds). From Joel On Software (2008). (https://www.joelonsoftware.com/2008/02/19/why-are-the-microsoft-office-file-formats-so-complicated-and-some-workarounds/). Mentions use
- Weaponized MS Office 97-2003 legacy/binary formats (doc, xls, ppt, ...) | from Decalage (http://decalage.info/file_formats_security/office). Has many useful links related to Microsoft binary formats, with an emphasis on security. No longer actively updated.
- Resources related to i4i v. Microsoft (2009, USA), a suit filed on March 8, 2007 alleging that Microsoft's Custom XML feature, when used in XML files, infringed its U.S. patent 5,787,449, issued July 28, 1998. Google's interface to the same patent data is at ht
  - Memorandum Opinion and Order: for i4i Ltd. Partnership v. Microsoft Corp., 670 F. Supp. 2d 568 - Dist. Court, ED Texas 2009 (https://scholar.google.com/scholar_case?case=11309803085850071839).
  - i4i v. Microsoft (2009, USA) | Notes from swpat.org (http://en.swpat.org/wiki/I4i_v._Microsoft_(2009,_USA)). Includes timeline related to case. Swpat.org supports campaigns against software patents
  - Microsoft Word Lawsuit: XML Explained (2009/08/13) | from PC World (https://www.pcworld.com/article/170148/microsoft_word_xml.html).
  - Microsoft loses its appeal in $200-million-plus Custom XML patent infringement case (2009/12/22) | from ZDNet (https://www.zdnet.com/article/microsoft-loses-its-appeal-in-200-million-plus-custom-xml-patent-infringement-case/).
  - What is "Custom XML?" ... and the impact of the i4i judgment on Word (2009/12/22) | from Microsoft blog for developers (https://docs.microsoft.com/en-us/archive/blogs/gray_knowlton/what-is-custom-xml-and-the-impact-of-the-i4i-judgment-on-word). Incl
  - Using Content Controls vs. Custom XML Elements (2010/01/03) | from Microsoft blog (https://docs.microsoft.com/en-us/archive/blogs/gray_knowlton/using-content-controls-vs-custom-xml-elements).
  - Regarding Custom XML Patch distribution and availability (2010/01/10) | from Microsoft blog for developers (https://docs.microsoft.com/en-us/archive/blogs/gray_knowlton/regarding-custom-xml-patch-distribution-and-availability).
  - Scanning Tool to Detect Custom XML Markup in .DOCX and .DOCM files (2010/01/19) | from Microsoft blog (https://docs.microsoft.com/en-us/archive/blogs/gray_knowlton/scanning-tool-to-detect-custom-xml-markup-in-docx-and-docm-files). Microsoft di
  - Custom XML markup is removed when you open a document in Word 2013 | from Microsoft documentation (https://docs.microsoft.com/en-us/office/troubleshoot/word/custom-xml-markup-is-removed). Includes summary result of the court ruling and reco
  - Content controls in Word | from Microsoft documentation (https://docs.microsoft.com/en-us/office/client-developer/word/content-controls-in-word).
- The PRONOM database from the U.K. National Archives has a number of records for various versions of format from Microsoft with extension .doc and based on the Compound File Binary (CFB) container format. The entry corresponding to the version desc
  - PRONOM entry for fmt/40 (http://www.nationalarchives.gov.uk/PRONOM/fmt/40). Information in PRONOM on Microsoft Word Document 97-2003. PUID: fmt/40
  - PRONOM entry for fmt/39 (http://www.nationalarchives.gov.uk/PRONOM/fmt/39). Information in PRONOM on Microsoft Word Document 6.0/95. PUID: fmt/39
  - PRONOM entry for fmt/609 (http://www.nationalarchives.gov.uk/PRONOM/fmt/609). Information in PRONOM on Microsoft Word (Generic) 6.0-2003. PUID: fmt/609
- Wikidata entry for Q28858035 (https://www.wikidata.org/wiki/Q28858035). Information in Wikidata on Word Binary File Format, version nFib=0x00C1. WikiData Title ID: Q28858035
- Wikidata entry for Q888498 (https://www.wikidata.org/wiki/Q888498). Information in Wikidata on Word Binary File Format, all versions. WikiData Title ID Q888498

* format specifications

* * useful references



Main | Table of Contents | Introduction | Textual Works | Still Image Works | Moving Image Works | Audio Works | Musical Scores | Datasets | GIS, Geospatial and Non-GIS Cartographic | Design and 3D | Software and Video Games | Web Archives

Library of Congress Recommended Formats Statement - 2020-2021

Recommended Formats Statement identifies hierarchies of the physical and technical characteristics of creative formats, both analog and digital, which will best meet the needs of all concerned, maximizing the chances for survival and continued accessibility of creative content well into the future.

The 2020-2021 version includes significant changes from the 2019-2020 version. Specific changes are detailed in the Change Log and the Introduction.

Recommended Formats Statement

https://www.loc.gov/preservation/resources/rfs/index.html

content categories

textual works | still image works | moving image works | audio works | musical scores | datasets | GIS, geospatial and non-GIS cartographic | design and 3D | software and video games | web archives

orange = new or changed content categories for 2020

evaluation criteria

Global/Community Format Sustainability Factors

Disclosure
Adoption
Transparency
Self-documentation
External dependencies
Impact of patents
Technical protection mechanisms

LC Local/Institutional Factors

Staff experience and expertise

Software/Hardware/OS available

Representation/extent in LC collections/storage

Established workflow/functionality

PREFERRED

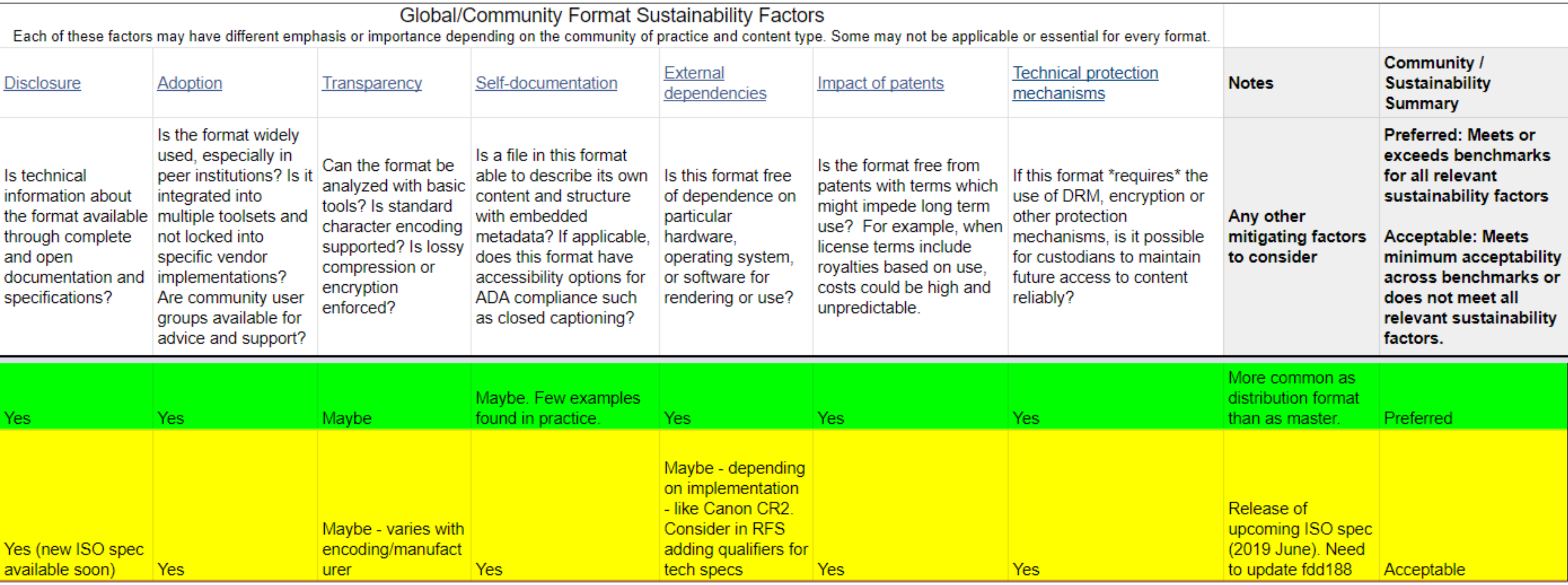
Global/community: Meets or exceeds benchmarks for all relevant sustainability factors

Local/institutional: The Library of Congress has the skills, experience, workflows, tools and systems to manage and preserve these formats in current systems with confidence.

ACCEPTABLE

Global/community: Meets minimum acceptability across benchmarks or does not meet all relevant sustainability factors.

Local/institutional: The Library of Congress can manage this format at a basic level of acquisition, management and preservation; and a greater ability for management and preservation is within the Library's capacity with further investment.



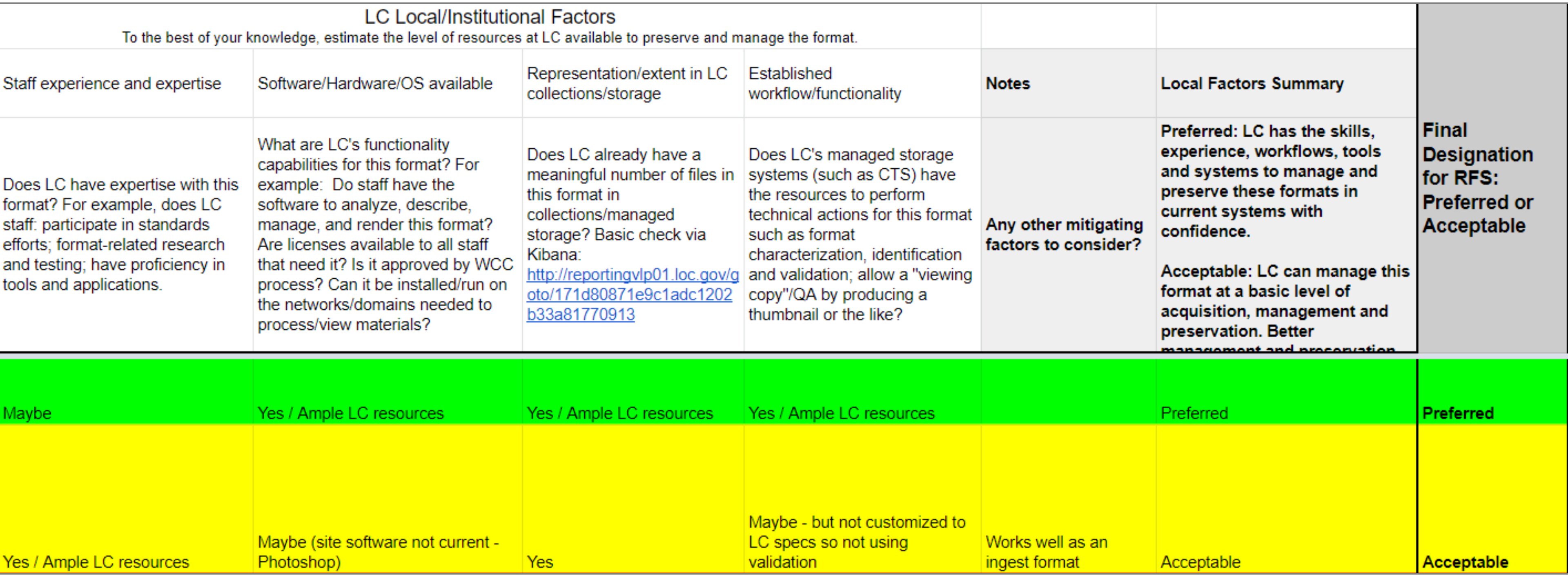
Global/Community Format Sustainability Factors

Each of these factors may have different emphasis or importance depending on the community of practice and content type. Some may not be applicable or essential for every format.

Disclosure	Adoption	Transparency	Self-documentation	External dependencies	Impact of patents	Technical protection mechanisms	Notes	Community / Sustainability Summary
Is technical information about the format available through complete and open documentation and specifications?	Is the format widely used, especially in peer institutions? Is it integrated into multiple toolsets and not locked into specific vendor implementations? Are community user groups available for advice and support?	Can the format be analyzed with basic tools? Is standard character encoding supported? Is lossy compression or encryption enforced?	Is a file in this format able to describe its own content and structure with embedded metadata? If applicable, does this format have accessibility options for ADA compliance such as closed captioning?	Is this format free of dependence on particular hardware, operating system, or software for rendering or use?	Is the format free from patents with terms which might impede long term use? For example, when license terms include royalties based on use, costs could be high and unpredictable.	If this format *requires* the use of DRM, encryption or other protection mechanisms, is it possible for custodians to maintain future access to content reliably?	Any other mitigating factors to consider	**Preferred: Meets or exceeds benchmarks for all relevant sustainability factors** **Acceptable: Meets minimum acceptability across benchmarks or does not meet all relevant sustainability factors.**
Yes	Yes	Maybe	Maybe. Few examples found in practice.	Yes	Yes	Yes	More common as distribution format than as master.	Preferred
Yes (new ISO spec available soon)	Yes	Maybe - varies with encoding/manufacturer	Yes	Maybe - depending on implementation - like Canon CR2. Consider in RFS adding qualifiers for tech specs	Yes	Yes	Release of upcoming ISO spec (2019 June). Need to update fdd188	Acceptable

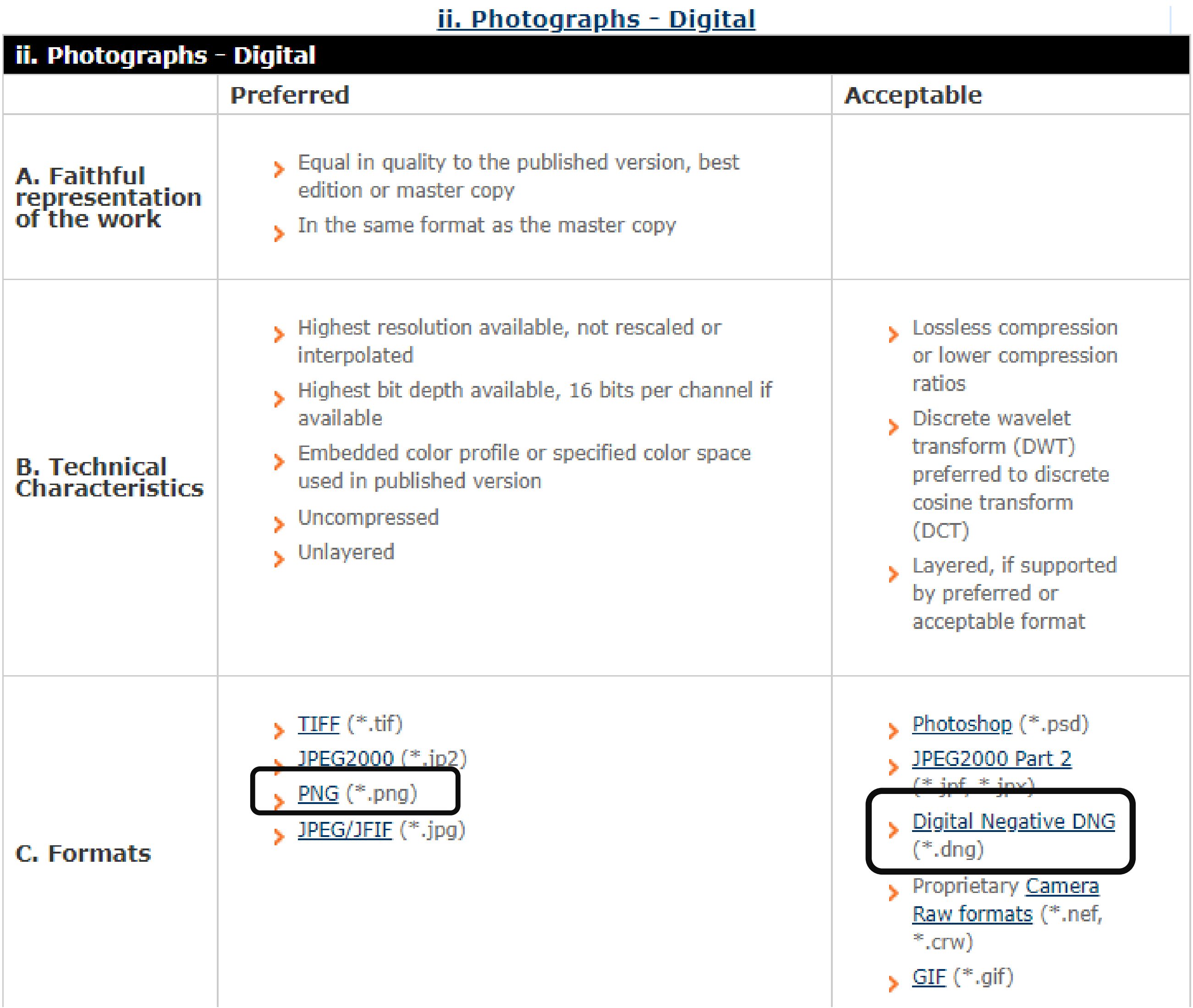
Green = PNG

Yellow = DNG



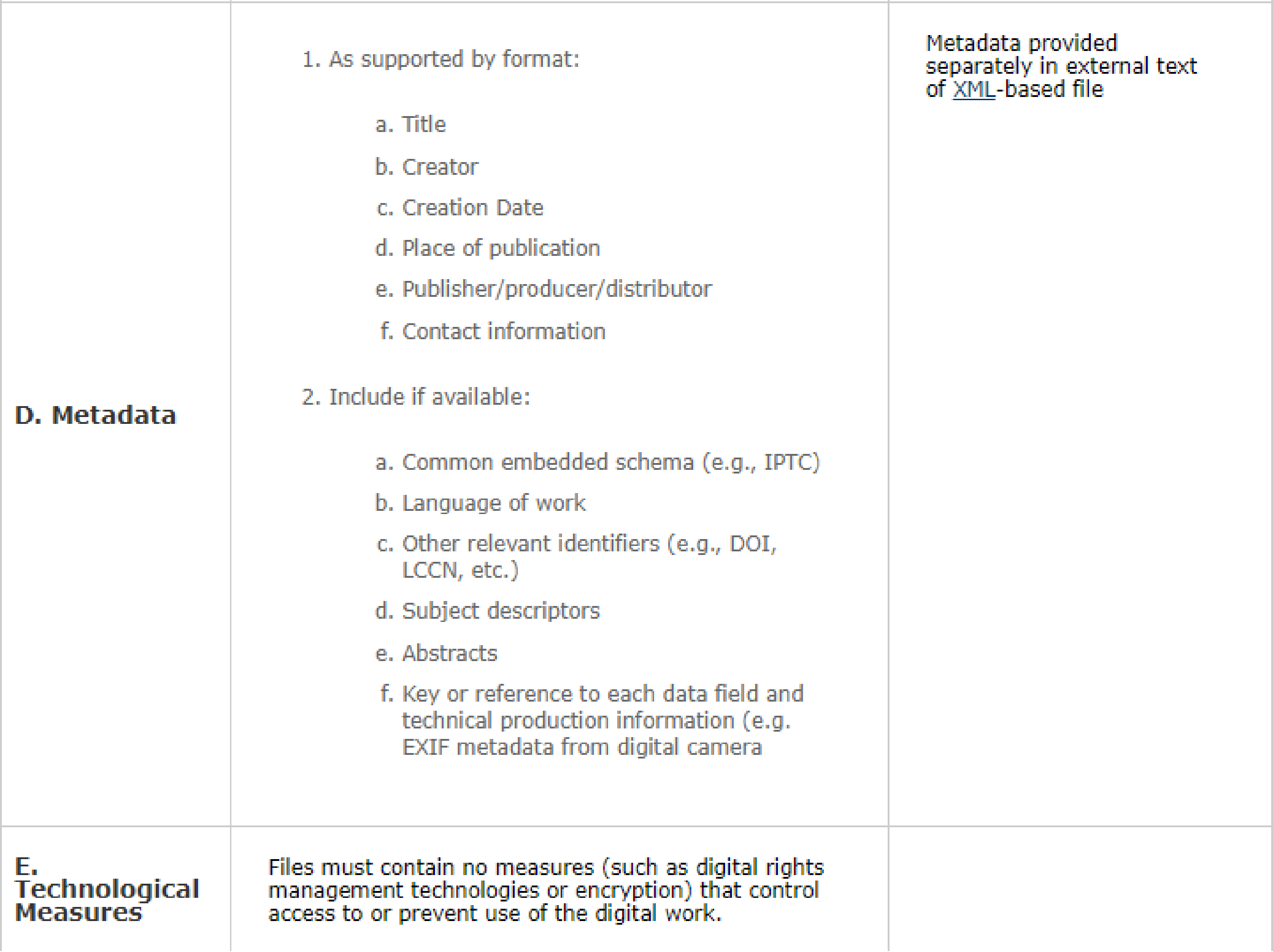
LC Local/Institutional Factors To the best of your knowledge, estimate the level of resources at LC available to preserve and manage the format.						**Final Designation for RFS: Preferred or Acceptable**
Staff experience and expertise	Software/Hardware/OS available	Representation/extent in LC collections/storage	Established workflow/functionality	**Notes**	**Local Factors Summary**	
Does LC have expertise with this format? For example, does LC staff: participate in standards efforts; format-related research and testing; have proficiency in tools and applications.	What are LC's functionality capabilities for this format? For example: Do staff have the software to analyze, describe, manage, and render this format? Are licenses available to all staff that need it? Is it approved by WCC process? Can it be installed/run on the networks/domains needed to process/view materials?	Does LC already have a meaningful number of files in this format in collections/managed storage? Basic check via Kibana: http://reportingvlp01.loc.gov/goto/171d80871e9c1adc1202b33a81770913	Does LC's managed storage systems (such as CTS) have the resources to perform technical actions for this format such as format characterization, identification and validation; allow a "viewing copy"/QA by producing a thumbnail or the like?	**Any other mitigating factors to consider?**	**Preferred: LC has the skills, experience, workflows, tools and systems to manage and preserve these formats in current systems with confidence.** **Acceptable: LC can manage this format at a basic level of acquisition, management and preservation. Better management and preservation**	
Maybe	Yes / Ample LC resources	Yes / Ample LC resources	Yes / Ample LC resources		Preferred	**Preferred**
Yes / Ample LC resources	Maybe (site software not current - Photoshop)	Yes	Maybe - but not customized to LC specs so not using validation	Works well as an ingest format	Acceptable	**Acceptable**

Green = PNG
Yellow = DNG



ii. Photographs - Digital

ii. Photographs - Digital	Preferred	Acceptable
A. Faithful representation of the work	Equal in quality to the published version, best edition or master copy In the same format as the master copy	
B. Technical Characteristics	Highest resolution available, not rescaled or interpolated Highest bit depth available, 16 bits per channel if available Embedded color profile or specified color space used in published version Uncompressed Unlayered	Lossless compression or lower compression ratios Discrete wavelet transform (DWT) preferred to discrete cosine transform (DCT) Layered, if supported by preferred or acceptable format
C. Formats	TIFF (*.tif) JPEG2000 (*.jp2) PNG (*.png) JPEG/JFIF (*.jpg)	Photoshop (*.psd) JPEG2000 Part 2 (*.jpf, *.jpx) Digital Negative DNG (*.dng) Proprietary Camera Raw formats (*.nef, *.crw) GIF (*.gif)



D. Metadata	1. As supported by format: a. Title b. Creator c. Creation Date d. Place of publication e. Publisher/producer/distributor f. Contact information 2. Include if available: a. Common embedded schema (e.g., IPTC) b. Language of work c. Other relevant identifiers (e.g., DOI, LCCN, etc.) d. Subject descriptors e. Abstracts f. Key or reference to each data field and technical production information (e.g. EXIF metadata from digital camera	Metadata provided separately in external text of XML-based file
E. Technological Measures	Files must contain no measures (such as digital rights management technologies or encryption) that control access to or prevent use of the digital work.	

https://www.loc.gov/preservation/resources/rfs/stillimg.html

acknowledgements

Sustainability of Digital Formats

https://www.loc.gov/preservation/digital/formats/index.html

- Caroline Arms
- Marcus Nappier
- Laurel Gassie

Recommended Formats Statement

https://www.loc.gov/preservation/resources/rfs/index.html

- Ted Westervelt
- Jesse Johnston
- Marcus Nappier
- All content team leaders/members

thank you

Kate Murray

kmur@loc.gov | @fileformatology