Library of Congress
Recommended Formats Statement
2017-2018

Introduction

Categories of creative content:

I. Textual Works and Musical Compositions
- i. Textual Works – Print
- ii. Textual Works – Digital
- iii. Textual works – Electronic Serials
- iv. Digital Musical Compositions (score-based representations)

II. Still Image Works
- i. Photographs – Print
- ii. Photographs – Digital
- iii. Other Graphic Images – Print
- iv. Other Graphic Images – Digital
- v. Microforms

III. Audio Works
- i. Audio – On Tangible Medium (digital and analog)
- ii. Audio – Media-independent (digital)

IV. Moving Image Works
- i. Motion Pictures – Digital and Physical Media
- ii. Video – File-Based and Physical Media

V. Software and Electronic Gaming and Learning

VI. Datasets/Databases
- i. Datasets
- ii. Geospatial Data
- iii. Databases

VII. Websites

Introduction to the 2017-2018 revision

Throughout its history, the Library of Congress has been committed to the goal best described in its mission statement to provide "the American people with a rich, diverse, and enduring source of knowledge that can be relied upon to inform, inspire, and engage them, and support their intellectual and creative endeavors". At its core, this source of knowledge is best expressed in the Library's unparalleled collection of creative works. The quality of the collection reflects the Library's care in selecting materials and the effort it invests in preserving them and making them accessible to the American people for the long term.

To build such a substantial and wide-ranging collection and to ensure that it will be available for successive generations, the Library relies upon many things. In order to maximize the scope and scale of the content in the collection, the Library calls upon the wealth of expertise in languages, subject matter and trends in publishing and content creation provided by the specialists who identify and acquire material for the Library's collection. But knowledge of the technical characteristics of the production of creative works is required as well. In the past, the lasting power of the collections depended exclusively upon the endurance of such materials as the paper, ink, and binding of a book; the acetate or paper coated with gelatin in a photograph; or the shellac, vinyl, and coated polyester that comprise a sound recording. Although these materials remain in use today, creators and publishers have also begun to employ a wide array of intangible digital formats, as well as continuing to change and adapt the physical formats in which they work. The Library needs to be able to identify the formats which are suitable for large-scale acquisition and preservation for long-term access if it is to continue to build its collection and ensure that it lasts into the future.

The Library is also well aware that it shares with many other institutions, organizations and corporations the common goal of the creation, dissemination and preservation of knowledge, which benefits the nation socially, economically and politically. And the Library knows that this common goal is partially expressed in the work of identifying the formats of creation best suited to preservation and long-term access which other institutions are also both supporting and carrying out. In creating a statement which lays out the formats, both analog and digital, which will best meet the needs of all concerned, maximizing the chances for survival and continued accessibility of this creative content well into the future, the Library seeks to offer the benefit of its knowledge and expertise as well as its rather unique perspective in pursuit of this common goal. And by sharing this with others striving to this common goal, it seeks to receive feedback from them, bringing their own knowledge, expertise and perspective to the benefit of this effort.

In 2014 the Library developed the Recommended Format Specifications (in 2015 renamed the Recommended Formats Statement), which encompass the formats which it has decided to recommend, both internally to its own professionals and externally to creators, vendors and archivists, as the preferred ones to use to ensure the preservation and long-term access of the creative output of the nation and the world. This was the result of a three-year-long process which called upon expert knowledge of the creative and publishing landscape and marketplace. The Library was able to identify six broad categories of creative output: Textual Works and Musical Compositions; Still Image Works; Audio Works; Moving Image Works; Software and Electronic Gaming and Learning; and Datasets/Databases. In 2016, the Library expanded the scope of the Statement to include a seventh, Websites, the inclusion of which did much to expand the comprehensiveness of the Statement. Within each of these broad categories, there are further subdivisions, down to specific recommendations on various aspects of print books, digital photographs, motion picture films or electronic datasets, for

example. With this framework, the Library intends to define a sufficient space within which creative works should have the flexibility to grow and develop, and also help ensure that these creative works be accessible and authentic into the future.

Given the dynamic nature of the creative world, the Library does not intend this framework to stand rigid. Every year there is an annual review of the Statement, in which the Library solicits feedback, both internally and externally and then sets aside a three-month period within which its teams of experts make whatever revisions they deem necessary to ensure that the Statement remains engaged with the world around it and as accurate as we can make it. This does not tend to result in root-and-branch changes as by engaging with the Statement on an annual basis, the revisions are smaller and more manageable as less time has elapsed for the Statement to slip out of sync with developments in the creative world.

The Recommended Formats Statement is not intended to serve as an answer to all the questions raised in preserving and providing long-term access to creative content. It does not provide instructions for receiving material into repositories, managing that content or undertaking the many ongoing tasks which will be necessary to maintain this content so that it may be used well into the future. Tackling each of those aspects is a project in and of itself as each form of content has a unique set of facets and nuances. The Statement provides guidance on identifying sets of formats which are not drawn so narrowly as to discourage creators from working within them, but will instead encourage creators to use them to produce works in formats which will make preserving them and making them accessible simpler. The Library hopes that the Statement will help make it realistic to build, grow and save creative output for our individual and collective benefit for generations to come.

At a time of such great growth in the production of creative output, not only are the frontiers expanding, but new ones seem to crop up faster than we can grasp them, there is a definite need for some expert guidance, so that this amazing creative content is not lost to us. The Library of Congress appreciates that it is uniquely positioned to help provide that guidance and, in fact, that its position has given it a responsibility to take a lead in this effort. The Library is the nation's premier institution entrusted with creating and managing a rich, diverse, and enduring source of knowledge which will support the intellectual and creative endeavors of the American people. In producing and publishing the Recommended Formats Statement, it seeks to meet that charge, and to provide the benefit of its expertise for creators, vendors, and archivists, so that they might succeed in their goals to share and disseminate their creative output and benefit the nation generally.

I. Textual Works and Musical Compositions

i. Textual Works – Print (*books, musical compositions, etc.)*		
	Preferred	**Acceptable**
A. Paper	1. Archival quality paper (ISO 11108: 1996 for Archival Paper)	
B. Printing Process, in order of preference	1. Lithography (offset printing press) 2. Electrophotography (digital press) 3. Inkjet (inkjet printer using stable pigment or dye-based inks)	
C. Binding and Packaging	1. Slip-cased, if available 2. Binding, in descending order of preference: a. Hard cover i. Library binding (NISO Z39.78-2000) ii. Sewn iii. Glued only b. Soft cover i. Sewn ii. Glued only iii. Spiral- or plastic-bound iv. Stapled c. Loose-leaf (including all binders and indexes published as part of the deposit and offered for sale and distribution)	
D. Size	1. Larger-sized editions (Note: large-type editions are not preferred over editions with conventional size typefaces) 2. For broadsides and musical compositions, the Library prefers items: a. In protective folders b. Rolled (rather than folded)	
E. Rarity, Special Features, Illustrations	1. Limited editions (including those with special binding or special features) 2. Editions with the greatest number of unique features (such as pop-ups, overlaps, magnifiers, overlays, tabs, notches, etc.) 3. Illustrated editions; original color illustrations preferred over black and white reproductions	
F. Completeness	1. Complete work. For items published in a finite number of separate components, all elements published as part of the work and offered for sale or distribution must be submitted.	

	2. All updates, supplements, releases, and supersessions published as part of the work and offered for sale or distribution must be submitted. Insertions (including all binders and indexes) must be received in a regular and timely manner for proper maintenance of the deposit. 3. For unaccompanied vocal musical compositions: open score, with each part on separate staff 4. For vocal musical compositions with orchestral accompaniment and for instrumental musical compositions: a. Full score and up to 13 parts, if applicable; if published only by rental, lease, or lending, submit full score only b. Conductor's score and up to 13 parts, if applicable; if published only by rental, lease, or lending, submit conductor's score only	
G. Metadata	1. As displayed on item: a. Title b. Creator c. Creation Date or Start Date/End Date d. Place of Publication e. Publisher/Producer/Distributor f. ISBN 2. As displayed on item, if available: a. Other relevant identifiers (e.g., DOI, LCCN, etc.) b. Edition c. Subject descriptors d. Abstracts	

ii. Textual Works – Digital (*Electronic books, etc. For electronic serials, see sec. I.iii, below; for digital musical compositions, see sec. I.iv, below)*		
	Preferred	**Acceptable**
A. Technical Characteristics, in order of preference	1. Character encoding, in descending order of preference: a. UTF-8, UTF-16 (with BOM), US-ASCII b. ISO 8859	1. Other character encodings not listed here
B. Formats, in order of preference	1. XML-based markup formats, with included or accessible DTD/schema, XSD/XSL presentation stylesheet(s), and explicitly stated character encoding a. BITS-compliant (NLM Book DTD) b. EPUB-compliant c. Other widely-used book DTDs/schemas (e.g., TEI, DocBook, etc.) 2. Page-layout formats a. PDF/UA (ISO 14289-1-compliant) b. PDF/A (ISO 19005-compliant) c. PDF (highest quality available, with features such as searchable text, embedded fonts, lossless compression, high resolution images, device-independent specification of colorspace, content tagging; includes document formats such as PDF/X)	1. Other structured or markup formats a. XHTML or HTML, with DOCTYPE declaration and presentation stylesheet(s) b. XML-based document formats (widely-used and publicly-documented), with presentation stylesheet(s) if applicable. Includes DOCX/OOXML 2012 (ISO 29500), ODF (ISO/IEC 26300) and OOXML (ISO/IEC 29500). c. SGML, with included or accessible DTD d. Other XML-based non-proprietary formats, with presentation stylesheet(s) e. XML-based formats that use proprietary DTDs or schemas, with presentation stylesheet(s) 2. Page-layout formats a. PDF (web-optimized) 3. Other formats a. Rich text format b. Plain text c. Widely-used proprietary word-processing formats d. Other text- or graphic-based formats not listed here that represent textual works

C. Rarity and Special Features	1. Limited editions (including those with special features such as high resolution images) 2. Editions with the greatest number of unique features (such as additional content, multimedia, interactive elements, etc.)	
D. Completeness	1. Complete work. For items published in a finite number of separate components, all elements published as part of the work and offered for sale or distribution must be submitted. Includes all associated external files and fonts considered integral to the publication. 2. All updates, supplements, releases, and supersessions published as part of the work and offered for sale or distribution must be submitted and received in a regular and timely manner for proper maintenance of the deposit.	
E. Metadata	1. As supported by format (e.g., standards-based formats such as ONIX, XMP, MODS, or MARCXML either embedded in or accompanying the digital item): a. Title b. Creator c. Creation Date or Start Date/End Date d. Place of publication e. Publisher/ producer/ distributor f. ISBN g. Contact information 2. Include if available: a. Language of work b. Other relevant identifiers (e.g., DOI, LCCN, original URL, etc.) c. Edition d. Subject descriptors e. Abstracts	
F. Technological Measures	1. Files must contain no measures (such as digital rights management or encryption) that control access to or prevent use of the digital work.	

iii. Textual Works – Electronic serials		
	Preferred	**Acceptable**
A. Technical Characteristics, in order of preference	1. Character encoding, in descending order of preference: a. UTF-8, UTF-16 (with BOM), US-ASCII b. ISO 8859	1. Other character encodings not listed here
B. Formats, in order of preference	1. Content compliant with the NISO JATS: Journal Article Tag Suite (ANSI/NISO Z39.96-2015) with XSD/XSL presentation stylesheet(s) and explicitly stated character encoding 2. Page-layout formats a. PDF/UA (ISO 14289-1-compliant) b. PDF/A (ISO 19005-compliant) c. PDF (highest quality available, with features such as searchable text, embedded fonts, lossless compression, high resolution images, device-independent specification of colorspace; content tagging; includes document formats such as PDF/X)	1. Other structured or markup formats: a. Widely-used serials or journal non-proprietary XML-based DTDs/schemas with included or accessible DTD/schema, presentation stylesheet(s) and explicitly stated character encoding. b. Proprietary XML-based format for serials or journals (with documentation) with DTD/schema and presentation stylesheet(s) c. XHTML or HTML, with DOCTYPE declaration and presentation stylesheet(s) d. XML-based document formats (widely used and publicly documented). With presentation stylesheets, if applicable. Includes DOCX/OOXML 2012 (ISO 29500), ODF (ISO/IEC 26300) and OOXML (ISO/IEC 29500). 2. Page-layout formats a. PDF (web-optimized with searchable text) 3. Other formats a. Rich text format b. Plain text c. Widely-used proprietary word processing or page-layout formats

		d. Other text- or graphic-based formats not listed here that represent textual works
C. Completeness	1. Complete work. All elements considered integral to the publication and offered for sale or distribution must be submitted – e.g., articles, table(s) of contents, front matter, back matter, etc. Includes all associated external files and fonts considered integral to the publication. 2. All updates, supplements, releases, and supersessions published as part of the work and offered for sale or distribution must be submitted and received in a regular and timely manner for proper maintenance of the deposit.	
D. Metadata	1. Title-level metadata (e.g., standards-based formats such as ONIX, XMP, MODS, or MARCXML either embedded in or accompanying the digital item): a. Serial or journal title b. ISSN and ISSN-L c. Publisher d. Frequency e. Place of publication 2. Article-level metadata as relevant or applicable (e.g., standards-based formats such as ONIX, XMP, MODS, or MARCXML either embedded in or accompanying the digital item): a. Volume(s) b. Number(s) c. Issue date(s) d. Article title(s) e. Article author(s) f. Article identifier (DOI, original URL, etc.) 3. Include if available: a. Other descriptive metadata (e.g., subject heading(s), descriptor(s), abstract(s))	
E. Technological Measures	1. Files must contain no measures (such as digital rights management or encryption) that control access to or prevent use of the digital work.	

iv. Digital Musical Compositions (score-based representations)		
	Preferred	**Acceptable**
A. Technical Characteristics, in order of preference	1. Character Encoding, in descending order of preference: a. UTF-8, UTF-16 (with BOM), US-ASCII b. ISO 8859	1. Other character encodings not listed here
B. Formats, in order of preference	1. XML-based markup music notational format, with included or accessible DTD/schema, XSD/XSL presentation stylesheet(s), and explicitly stated character encoding. a. MusicXML b. Music Encoding Initiative (MEI) c. Other widely-used and publicly documented musical notation DTDs/schemas 2. Page-layout formats a. PDF/UA (ISO 14289-1-compliant) b. PDF/A (ISO 19005-compliant) c. PDF (highest quality available, with features such as searchable text, embedded fonts, lossless compression, high resolution images; includes document formats such as PDF/X)	1. Other structured or markup formats a. XHTML or HTML, with DOCTYPE declaration and presentation stylesheet(s) b. SGML, with included or accessible DTD c. XML-based musical notation formats that use proprietary DTDs or schemas, with presentation stylesheet(s) 2. Page-layout formats a. PDF (web-optimized) 3. Other formats a. Widely-used proprietary music notation formats b. Other music composition formats (including graphics-based formats) not listed here
C. Rarity and Special Features	1. Limited editions (including those with special features) 2. Editions with the greatest number of unique features (such as additional content, multimedia, interactive elements, etc.)	
D. Completeness	1. Complete work. For items published in a finite number of separate components, all elements published as part of the work and offered for sale or distribution must be submitted. Includes all associated external	

	files and fonts considered integral to the publication. 2. All updates, supplements, releases, and supersessions published as part of the work and offered for sale or distribution must be submitted and received in a regular and timely manner for proper maintenance of the deposit. 3. For unaccompanied vocal musical compositions: open score, with each part on separate staff 4. For vocal musical compositions with orchestral accompaniment and for instrumental musical compositions: a. Full score and all parts, if applicable; if published only by rental, lease, or lending, full score only may be submitted b. Conductor's score and all parts, if applicable; if published only by rental, lease, or lending, conductor's score only may be submitted	
E. Metadata	1. As supported by format (e.g., standards-based formats such as ONIX, XMP, MODS, or MARCXML either embedded in or accompanying the digital item): a. Title b. Creator c. Creation Date or Start Date/End Date d. Place of publication e. Publisher/ producer/ distributor f. ISMN g. Instrumentation 2. Include if available: a. Language of work b. Other relevant identifiers (e.g., ISBN, DOI, LCCN, original URL, etc.) c. Edition	

	d. Subject descriptors e. Event f. Abstracts	
F. Technological Measures	1. Files must contain no measures (such as digital rights management or encryption) that control access to or prevent use of the digital work.	

II. Still Image Works

i. Photographs – Print		
	Preferred	**Acceptable**
A. Faithful representation of the work	1. Equal in quality to the publication version or master copy	
B. Permanence and appearance	1. Unmounted 2. Pigmented inks (if digitally printed) 3. Fixed, well-washed (if wet chemistry method)	
C. Size	1. Min: 8 x10” 2. Max: 28 x 36” 3. If multiple sizes, largest available	
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. Language of work b. Other relevant identifiers (e.g., DOI, LCCN, etc.) c. Subject descriptors d. Abstracts e. Key or reference to each data field and technical production information (type of paper, how processed, publisher internal tracking numbers)	

ii. Photographs – Digital		
	Preferred	**Acceptable**
A. Faithful representation of the work	1. Equal in quality to the published version or master copy 2. In the same format as the master copy	
B. Technical Characteristics	1. Highest resolution available, not rescaled or interpolated 2. Highest bit depth available, 16 bits per channel if available 3. Embedded color profile or specified color space used in published version 4. Uncompressed 5. Unlayered	1. Lossless compression or lower compression ratios 2. Discrete wavelet transform (DWT) preferred to discrete cosine transform (DCT) 3. Layered, if supported by preferred or acceptable format
C. Formats	1. TIFF (*.tif) 2. JPEG2000 (*.jp2) 3. PNG (*.png) 4. JPEG/JFIF (*.jpg) 5. Digital Negative DNG (*.dng) 6. BMP (*.bmp) 7. GIF (*.gif)	1. Photoshop (*.psd, *.psb) 2. JPEG2000 Part 2 (*.jpf, *.jpx) 3. Proprietary Camera Raw formats (*.nef, *.crw)
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, FGDC, ISO 19115) b. Language of work c. Other relevant identifiers (e.g., PLUS ID, 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)	1. Metadata provided separately in external text of xml-based file
E. Technological Measures	1. Files must contain no measures (such as digital rights management or encryption) that control access to or prevent use of the digital work.	

iii. Other Graphic Images – Print (*e.g., posters, architectural drawings, postcards, maps, fine prints*)		
	Preferred	**Acceptable**
A. Faithful representation of the work	1. Equal in quality to the publication version or master copy	
B. Permanence and appearance	1. Packaging materials equivalent to published form (e.g., binding, box/packaging materials) 2. If multiple versions available, provide the most widely distributed edition. 3. If limited edition, provide an unnumbered but otherwise identical copy. 4. For large items, provide rolled, unfolded. 5. Prefer edition with protective coatings. 6. Faithful representation of the work as published. Equal quality to publication version.	
C. Cartographic materials, in order of preference	1. Most complete data (all layers, appendices) 2. Largest size 3. Most widely distributed	
D. Related Materials	1. Includes indexes, study guides or other matter if available 2. Also includes annotations, accompanying tabular or textual matter or other interpretive aids	
E. 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. Language of work b. Other relevant identifiers (e.g., DOI, LCCN, etc.) c. Subject descriptors d. Abstracts e. Key or reference to each data field and technical production information (type of paper, how processed, publisher internal tracking numbers)	

iv. Other Graphic Images – Digital		
	Preferred	**Acceptable**
A. Faithful representation of the work	1. Equal in quality to the published version or master copy 2. In the same format as the master copy	
B. Technical Characteristics	1. Highest resolution available, not rescaled or interpolated 2. Highest bit depth available, 16 bits per channel if available 3. Specified color space used in published version 4. Uncompressed 5. Unlayered	1. Lower compression ratios 2. Discrete wavelet transform (DWT) preferred to discrete cosine transform (DCT) 3. Layered, if supported by preferred or acceptable format
C. Formats (raster)	1. TIFF (*.tif) 2. JPEG2000 (*.jp2) 3. PNG (*.png) 4. JPEG/JFIF (*.jpg) 5. Digital Negative DNG (*.dng) 6. BMP (*.bmp) 7. GIF (*.gif)	1. Photoshop (*.psd, *.psb) 2. JPEG2000 Part 2 (*.jpf, *.jpx) 3. MrSID (*.sid) 4. Encapsulated Postscript (*.eps) 5. Proprietary Camera Raw formats
D. Formats (vector)	1. Scalable vector graphics (*.svg) 2. AutoCAD Drawing Interchange Format (*.dxf) 3. Encapsulated Postscript (*.eps) 4. Shapefile	1. Computer Graphics Metafile (CGM, WebCGM) 2. Non-proprietary formats endorsed as standards by a professional community or government agency, e.g. IFC, STEP 3. Page-layout formats, e.g. PDF/UA (ISO 14289-1-compliant), PDF/A (ISO 19005-compliant), PDF (highest quality available, with features such as searchable text, embedded fonts, lossless compression, high resolution images; includes document formats such as PDF/X) 4. Proprietary vector formats
E. Cartographic materials, in order of preference	1. Most complete data (all layers, appendices), even if proprietary 2. GeoTIFF 3. GeoPDF	

(for geospatial data, see sec. VI.ii)	4. GeoJPEG2000 5. Shapefile	
F. Related Materials	1. Includes indexes, study guides or other matter if available 2. Also includes annotations, accompanying tabular or textual matter or other interpretive aids	
G. 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., FGDC, ISO 19115) 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)	1. Metadata provided separately in external text or XML-based file
H. Technological Measures	1. Files must contain no measures (such as digital rights management or encryption) that control access to or prevent use of the digital work.	

v. Microforms		
	Preferred	**Acceptable**
A. Faithful representation of the work	1. Equal in quality to the publication version or master copy	
B. Permanence and appearance	1. Silver halide 2. Positive polarity 3. Color (when available) 4. Polyester film base	
C. Format (newspapers and newspaper-formatted serials)	1. Roll microfilm	
D. Format (all other materials), in order of preference	1. Microfiche 2. Roll microfilm 3. Microfilm cassettes 4. Micro-opaque prints	
E. Size	1. 35mm, if roll film	1. Other sizes that match the primary production master
F. Related Materials	1. Include indexes, study guides or other printed matter if available	
G. Metadata	3. As supported by format a. Title b. Creator c. Creation Date d. Place of Publication e. Publisher/producer/distributor f. Contact Information 4. Include if available: a. Language of work b. Other relevant identifiers (e.g., DOI, LCCN, etc.) c. Subject descriptors d. Abstracts e. Key or reference to each data field and technical production information (type of paper, how processed, publisher internal tracking numbers)	

III. Audio Works

i. Audio – On Tangible Medium (digital and analog)		
	Preferred	**Acceptable**
A. Sound Recordings, in order of preference	1. Final production/release version of content rather than pre-production version 2. Published Compact Disc (CD audio) rather than Recordable Compact Disc (CD-R, audio format) 3. With all jackets, sleeves, enclosures and inserts rather than without 4. Stereophonic if originally recorded/released as stereophonic 5. Monaural if originally recorded/released as monaural 6. Direct Stream Digital (DSD) or other multi-channel (e.g. Surround Sound) version in addition to stereophonic version if released in both 7. Vinyl disc (LP) in addition to Compact Disc (CD) if released in both	1. Recordable Compact Disc (CD-R) rather than vinyl disc 2. Vinyl disc rather than audio cassette 3. Audio cassette if only released as such

ii. Audio – Media-independent (digital)		
	Preferred	**Acceptable**
A. Audio, in order of preference	1. Final production /release version of content rather than pre-production version 2. Highest native resolution PCM WAVE file of final version produced (44.1 kHz / 16 bit or higher) in addition to Compact Disc (CD) when both are produced 3. WAVE file with embedded metadata (Broadcast WAVE) rather than without (LC will specify fields) 4. File in native resolution rather than up-sampled resolution 5. Very high resolution file formats such as DSD, PCM 176.4khz , 192khz up to 384kh when produced for release in addition to Compact Disc (CD) when both are produced 6. DSD in the released version (e.g., surround-sound or stereo)	1. Uncompressed file of final release version 2. Highest resolution compressed version in a major standard compression scheme 3. Lossless compression scheme rather than lossy compression scheme
B. Streaming/Podcasts, in order of preference	1. Uncompressed files rather than compressed. 2. Compressed version in a major standard compression scheme rather than non-standard scheme	
C. Accompanying Image/Text Files, in order of preference	1. With final version of all accompanying image and text files; higher resolution images rather than lower a. TIFF or JPEG formats for images b. Text files in PDF	1. TIFF, JPEG or PDF of text files of final release version 2. Other standard images and text formats
D. Metadata	1. Provide most complete metadata set as delivered to online distributors (e.g. iTunes and Amazon), which may include elements not embedded in a file, including but not limited to: a. Song/work title b. Album title c. Artist d. Composer e. Genre f. Publisher/label name and issue number g. Location and date of performance h. Date of publication i. Standard identifier (e.g. ISRC, UPC) j. Any other entity identifiers 2. Provide data in a standard XML-based format, such as the Electronic Release Notification (ERN-DDEX)	

E. Technological Measures	1. Files must contain no measures (such as digital rights management or encryption) that control access to or prevent use of the digital work.	

IV. Moving Image Works

i. Motion Pictures – Digital and Physical Media		
	Preferred	**Acceptable**
A. Theatrically Released Films, in order of preference	1. Complete final production/release version of motion picture work in the original production resolution, aspect ratio and frame rate 2. Theatrical release version in original gauge (e.g., 70mm, 35mm, 16mm) 3. Unencrypted interop Digital Cinema Package (DCP), when theatrical release is not distributed as film	1. Blu-ray disc, Blu-ray disc recordable
B. Audio	1. Complete final tracks, including any foreign language tracks and descriptive audio, when applicable	1. Each language and mix for the final production version shall be in its original channel structure and audio resolution as it was delivered to the content distributor
C. Metadata	1. Release title 2. Release/Production Date 3. Production Company and/or Producer 4. Distributor Name 5. Country of Origin 6. Language 7. Duration 8. Relevant unique identifiers applicable to the work (EIDR, ISAN)	
D. Technological Measures	1. Files must contain no measures (such as digital rights management or encryption) that control access to or prevent use of the digital work.	

ii. Videos – File-based and Physical Media		
	Preferred	**Acceptable**
A. Video – File-based , in order of preference	1. Final production version of content rather than pre-production version 2. Original production resolution and frame rate (i.e. 1080p24; 720p60, etc.) 3. Version and file-based format that was delivered to the content distributor	
B. Video – Physical Media, in order of preference		1. Final production version of content rather than pre-production version 2. Original production resolution and frame rate (i.e. 1080p24; 720p60, etc.) 3. Content contained in standard physical media in the following order of preference: a. HD: HDCAM-SR, HDCAM, HD-D5, Blu-ray disc, Blu-ray disc recordable b. SD: Digital Betacam, Betacam SP, DVD, DVD-r
C. Audio	1. Each language and mix for the final production version shall be in its original channel structure and audio resolution as it was delivered to the content distributor	
D. Metadata	1. Title (when applicable: Series Title, Season, Episode Title, Episode No., Production No.) 2. Content Type (e.g., Episode, Documentary, Oral History, News, Music Video, etc.) 3. Release/Broadcast/Production Date 4. Network 5. Production Company and/or Producer 6. Distributor Name 7. Country of Origin 8. Language 9. Duration 10. Relevant unique identifiers applicable to the work (EIDR, ISAN)	

E. Technological Measures	1. Files must contain no measures (such as digital rights management or encryption) that control access to or prevent use of the digital work.	

V. Software and Electronic Gaming and Learning

This category includes software for Desktop, Mobile/Handheld, Gaming and Learning Systems. **Note:** Gaming and Learning Systems are included in the software category, but often require submission of the proprietary platform as well (e.g. Game console) or a virtual or emulated version of the proprietary platform that runs on a commonly available operating system.

	Preferred	**Acceptable**
A. Content	1. With documentation and other accompanying material (e.g. instruction materials, errata, addenda, read me files). A copy of the electronic distribution file if the product was sold to the public as a downloadable file. Note: if the software was a part of a book publication (e.g. a software manual with accompanying discs), then a copy of the book must be submitted with the disc. 2. Source Code: A file-based copy of the source code from which the software was developed. Metadata that specifies which compiler was used to create the final code for commercial release—including the version number and build number of the compiler software—must be included. If the compiler is unique to the project or company (i.e. not commercially available), then a copy of compiler software in the specific version and build used to create this version of the software, along with specifications of the platform the compiler ran on, must be included in the submission. This submission should include the rights clearance for the Library to install and use the compiler to read and use the provided information. 3. Operating system: If not a commonly available OS, then a copy of the OS must be provided with the submission. **Note**: If there are different versions released at the same time for different operating systems (e.g. for Mac and Windows), a submission will be required for each. 4. Platform: If a submission requires a stand-alone or proprietary platform (e.g. a gaming system or child's toy), then a virtual or emulated version of the proprietary platform is required with submission.	1. Source code on physical media 2. Platform: If a submission requires a stand-alone or proprietary platform (e.g. a gaming system or child's toy), and a virtual or emulated version of the proprietary platform is not available, a physical platform is acceptable.

B. Delivery Method	1. Direct File Submission: These submissions would require grouping in a submission package such as BagIt, tar, or AXF object.	1. Mass storage device: All of the requested information may be included as distinct files or may be grouped together using a method such as BagIt, a tar, or AXF object a. Hard drive with USB (universal serial bus) interface b. CD-ROM disc c. Flash drive with USB interface d. DVD disc
C. Metadata	1. As supported by format: a. Title, b. Creator c. Creation date d. Place of publication e. Publisher/ producer/ distributor f. Contact information g. Production metadata such as credit, rights and files which are available at the time of production 2. Include if available: a. Language of work b. Other relevant identifiers (e.g., UPCDOI, LCCN, etc.) c. Subject descriptors d. Abstracts e. Key or reference to each data field	
D. Technological Measures	1. Files must contain no measures (such as digital rights management or encryption) that control access to or prevent use of the digital work.	1. A file containing digital rights management or encryption which requires a stand-alone or proprietary platform is accompanied by a virtual or emulated version of the proprietary platform or a physical version of the platform.

VI. Datasets/Databases

The Library is aware that, in some cases, the provision of datasets and databases for current research uses (including support for the U.S. Congress) may depend upon native formats and associated software, while preservation and long-term access may depend upon data-migration via transport or export formats, with a concomitant risk of loss of precision and accuracy. Given the focus of this document is preservation and long-term access, the following format preferences favor those outcomes.

i. Datasets (For Geospatial Data, see sec. VI.ii, below)		
	Preferred	**Acceptable, in order of preference**
A. Formats	1. Platform-independent, character-based formats are preferred over native or binary formats as long as data is complete, and retains full detail and precision. Preferred formats include well-developed, widely adopted, de facto marketplace standards, e.g. a. Self-describing, e.g. JSON, XML-based data formats using well known schemas, XML-based data formats accompanied by schema employed b. Line-oriented, e.g. TSV, CSV, fixed-width c. Platform-independent open formats, e.g. SQLite (.sqlite, .db, .db3) 2. Character Encoding, in descending order of preference: a. UTF-8, UTF-16 (with BOM), b. US-ASCII or ISO 8859-1 c. Other named encoding	1. Non-proprietary, publicly documented formats endorsed as standards by a professional community or government agency, e.g. CDF, HDF 2. Any proprietary format that is a de facto standard for a profession or supported by multiple tools (e.g. Excel .xls or .xlsx)
B. Related Materials	1. Consult the appropriate sections of this document to identify the preferred formats for supplementary material	
C. Delivery Method	1. Hard drive; CD-ROM; DVD-ROM, or other media not yet assigned	
D. Metadata	1. Deposits should include all applicable metadata, data dictionaries, XML schemas, and technical specifications as appropriate. Discipline-specific metadata standards should be used whenever possible 2. As supported by format: a. Title	

	b. Creator c. Creation date d. Place of publication e. Publisher/ producer/ distributor f. Contact information g. A list of software used to produce, render or compress the data (if applicable) h. Character encoding 3. Include if available: a. Language of work b. Other relevant identifiers (e.g., DOI, LCCN, canonical URL, etc.) c. Subject descriptors d. Abstracts e. Key or reference to each data field f. Checksums g. Version-control IDs or tags h. Information about how the data was collected and any sampling or post-processing which has been applied i. Known copyright terms, especially for datasets which combine data from multiple sources 4. For datasets serving as part of a database: proprietary database package and version	
E. Technological Measures	1. Files must contain no measures (such as digital rights management or encryption) that control access to or prevent use of the digital work. 2. Files in formats which support linking or embedding external resources (e.g. XML, JSON, Excel) should be self-contained to remain useful in the event of external service changes. 3. Files in formats which support executable code (e.g. Excel) do not contain executable code.	1. Files in formats which support executable code do not depend on embedded programs for purposes other than display (e.g. search, filtering, etc.); the raw data is available without executing code.

ii. Geospatial Data		
	Preferred	**Acceptable**
A. Formats, in order of preference	1. Most complete data (all layers, appendices), even if proprietary 2. Formats compatible with widely adopted GIS (e.g. ArcGIS) 3. Formats compatible with recommendations and tools from geospatial open source and open data communities a. Formats developed or endorsed by the Open Geospatial Consortium (OGC) (e.g., GML see: http://www.opengeospatial.org/) b. Formats supported by well supported open source software libraries such as GDAL, OGR and GeoTools	
B. Related Materials	1. Consult the appropriate sections of this document to identify the preferred formats for supplementary material	
C. Delivery Method	1. Hard drive; CD-ROM; DVD-ROM, or other media not yet assigned	
D. Metadata	1. For metadata information see Federal Geographic Data Committee (FGDC) 2. To the extent allowed by the underlying format, include available information about how the data was collected and any post-processing which has been applied	
E. Technological Measures	1. Files must contain no measures (such as digital rights management or encryption) that control access to or prevent use of the digital work.	

iii. Databases		
	Preferred	**Acceptable**
A. Preservation	1. Complete set of the content contained within the database, conforming to preferred specifications in sec. VI.i-ii	
B. Access, in order of preference	1. Publisher web interface with a. Comprehensive and user-friendly search and discovery b. Counter compliant usage statistics 2. Delivered preservation content	1. Documented API

VII. Websites

This format specification covers the Library's preferred format for archived web content, as well as a preferred "format" for presentation of web content for archiving (in other words, best practices for content creators to help in creating preservation-friendly websites). The Library is aware that websites, including blogs, social media and other web content that make up websites, are presented and created in formats for viewing in a web browser, and are often different than the standard format that is recommended for preservation and long-term access. Given that the focus of this document is preservation and long-term access, the following format preferences favor those outcomes, and include recommendations for best practices to better enable preservation of web content.

	Preferred	Acceptable
A. Technical Characteristics	1. Website creators can improve the archivability of web content by following best practices such as: a. Using sitemaps and stable URLs b. Using open formats c. Following accessibility standards, such as: i. Section 508 (https://www.access-board.gov/guidelines-and-standards/communications-and-it/about-the-section-508-standards/guide-to-the-section-508-standards) ii. Web Content Accessibility Guidelines (WCAG) (https://www.w3.org/WAI/intro/wcag) iii. United States Web Design Standards (https://standards.usa.gov/) d. Providing page specific titles and description, publication or update dates, and meaningful web addresses, when possible, to convey the substance of content presented 2. Resources that address this further and may be helpful to content creators can be found on the Library of Congress Guide to Creating Preservable Websites (http://loc.gov/webarchiving/preservable.html) a.	
B. Formats	1. The Library, and other organizations involved in web archiving, are preserving web content in the Web ARChive (WARC) format	2. Internet Archive's ARC_IA format, a precursor to the WARC format

C. Delivery Method	1. Capture using tools that produce non-proprietary output, to conform with standard formats and requirements	1. Transmission of WARC or ARC IA files created by web content producers or other archiving organizations
D. Metadata	1. Refer to the WARC ISO-standard specification for mandatory and recommended metadata fields 2. When displaying archived content, the following should be clearly indicated: a. archiving institution, b. dates and time of capture, c. statements about functionality within the archive to distinguish from the live site	1. The ARC IA should be named in a manner that easily identifies the archiving institution (see WARC standard for recommended naming conventions)
E. Technological Measures	1. Websites should not contain measures (such as content behind logins or only accessible through search functions) that control access to or prevent capture of the digital work. 2. Robots.txt restrictions should be set so as not to block crawlers from capturing important content, such as image and style sheets, which allow for replay of the site as it looked at the time of capture.	1. Tools currently available cannot capture all web content, so certain types of web content may not be preservable through web capture at this time. These include: a. Multi-media rich content b. Streaming media c. Deep web content d. Databases