



Library and Archives Canada

Guidelines on File Formats for Transferring Information Resources of Enduring Value



Bibliothèque et Archives
Canada

Library and Archives
Canada

Canada

Library and Archives Canada
**Guidelines on File Formats for
Transferring Information Resources of Enduring Value**

1. Effective Date

These Guidelines have been approved by the Senior Director General Innovation and Chief Information Officer Branch and take effect on October 1, 2014.

2. Application

These Guidelines provide advice on the digital file formats to be used when transferring information resources of enduring value (IREV) to Library and Archives Canada (LAC).

These Guidelines apply to all persons and organizations transferring digital IREV to LAC (hereinafter referred to as the “donor”).

These Guidelines supersede the *Local Digital Format Registry File Format Guidelines for Preservation and Long-term Access Version 1.0 (2010)*.

3. Definitions

See Appendix A.

4. Context

These Guidelines are part of LAC’s *Stewardship Policy Framework (2013)* and the accompanying *Policy on Holdings Management (July 2014)*. These documents mandate that IREV acquired and managed by LAC be accessible over time, and that consideration be given to stewardship requirements and resource capacity. The sustainability of IREV is therefore to be considered as part of all acquisition, stewardship, and reappraisal activities.

In accordance with sections 8 (2), and 10 of the *Library and Archives of Canada Act*, and section 2 (a) and (b) of the *Legal Deposit of Publications Regulations*, these Guidelines outline the appropriate file formats for submission to LAC of digital publications affected by legal deposit. While the *Library and Archives of Canada Act* section 10 (4) entitles LAC to collect all published versions and formats of a given title, LAC currently prefers to acquire publications in digital file formats defined within these Guidelines.

In accordance with sections 7, 12 and 13 of the *Library and Archives of Canada Act*, these Guidelines outline the appropriate digital formats that support any agreements between LAC and federal institutions for the transfer of digital IREV. Where such a transfer is governed by an existing records transfer agreement that specifies a digital format other than what is outlined in these Guidelines, federal institutions must consult with LAC prior to preparing the transfer.

These Guidelines will also apply to other acquisition agreements in which LAC representatives specify the file formats for the transfer of IREV.

5. Objective

These Guidelines restrict the number and types of file formats submitted to those formats that LAC has reasonable confidence can be preserved and made accessible over time, thereby ensuring sustainability.

6. Expected Results

Adherence to these Guidelines will allow LAC to achieve the following:

- Collaboration with donors on the long-term management and preservation of IREV;
- Acquisition of only those digital file formats identified as being sustainable;
- Transfer of digital IREV in a consistent, transparent and reliable manner that enables overall accountability;
- Alignment with international best practice in digital preservation.

7. Approach

File formats are specific patterns or structures that organize and define data. Some formats contain only one stream of uncompressed data, others may contain codecs to encode and compress the data and others may support several streams of media.

In addition to file formats, there are also container or encapsulating formats. These formats can contain and support various types or layers of data and metadata. Each of these formats may be handled by different programs, processes, or hardware but for the data stream to be interpreted properly, the information must be wrapped together.

The ability to preserve and use digital information is at risk if the computer hardware and software needed to access the information are no longer available or if the format specifications are not obtainable. The

use of appropriate file formats is therefore critical to sustainable long-term preservation. Due to a mix of technical and practical issues, certain file formats are more suitable for preservation.

The file format recommendations in these Guidelines are based on LAC's experience in collecting and preserving digital content as well as international best practices¹. In developing these Guidelines, LAC has attempted to balance the requirements for quality, stability, potential longevity and industry acceptance. Where possible, a preference has been placed on the selection of non-proprietary national and international standards, or failing this, on de facto industry standard file formats. De facto formats are widely used and recognized formats that have become industry standards because of their ubiquitous use and support and not because they have been formally approved by a standards organization. In some cases, LAC has also selected formats that it believes will become widely adopted in the near future.

The following criteria were considered when evaluating the sustainability of a given format:

Openness/transparency

- The relative ease with which knowledge of the file format and its technical information can be accumulated.

Adoption as a preservation standard

- The extent to which the format has been formally adopted by national libraries, archives and other memory institutions internationally.

Stability/compatibility

- The degree to which the format is backward and forward compatible.
- The degree to which the format is protected against file corruption.
- The relative frequency of updated or replacement versions of the format over time.

Dependencies/interoperability

- The degree to which the format relies on a particular hardware or software.

8. Scope

These Guidelines identify broad content categories covering all digital IREV acquired by LAC and provide transfer file format recommendations

¹ See Appendix B: Bibliography.

for each category. The file formats covered in this document have been divided into the following content categories² and subcategories:

- Text
- Presentations
- Email
- Still images
 - Digital photographs
 - Scanned text
- Digital audio
- Digital moving images
 - Digital cinema
 - Digital video
- Geospatial
- Computer Aided Design
- Data sets

The transfer file formats are identified as either:

- Preferred for transfer; or
- Acceptable for transfer.

Preferred formats are those formats that are readily usable and have been identified by LAC as possessing a high degree of long-term sustainability. These formats require little or no immediate management to achieve appropriate levels of preservation.

Acceptable formats are those that meet LAC's minimum criteria for sustainability. These formats may require LAC to perform some preservation actions on ingest to ensure their long-term sustainability.

All other formats are considered unacceptable because they do not meet the minimum requirements to be considered sustainable by LAC.

As a general rule, LAC will only accept file formats listed in these Guidelines. The onus is on the donor to ensure that IREV are in a preferred or acceptable file format. LAC reserves the right to refuse any file that is not in a preferred or acceptable file format and to request the migration of the files to a preferred or acceptable format. IREV may be exempted from compliance on a case-by-case basis after consultation

² Web content is not currently a content category because LAC actively harvests the web content that it seeks to acquire and preserve. Normally, LAC does not accept pre-harvested web content from donors. Any transfer of web content has to be negotiated with LAC.

with LAC representatives from the functional area responsible for acquisition.

These Guidelines do not contain information on creation, migration and capture standards. See LAC's *Standards on Digitization* (in development) for information on the production of digital IREV.

These Guidelines do not give information on the generation of metadata during the record creation process. See LAC's *Standards on Metadata* (in development).

These Guidelines do not outline how to achieve the actual physical or electronic transfer of IREV. Discuss the logistics of the transfer with the LAC representative responsible for the transfer.³

9. Transfer Requirements

When transferring digital IREV, identify the applicable content category and submit the resources in a preferred or acceptable format. Formats are listed by name and include a reference to the relevant specification that defines appropriate encoding methods. The formats in each section are organized alphabetically and do not imply an order of preference for any given format. However, LAC always prefers to receive a preferred file format over an acceptable file format.

Where required, the format category tables include a column that specifies the codec that must be used with each format. Donors must submit files that comply with both the format and codec that are listed.

In some cases, the donor must take additional steps to ensure that files are acceptable for long-term preservation by:

- Deactivating file level encryption;
- Deactivating digital rights management technologies;⁴
- Embedding in each record all fonts necessary to interpret the information;⁵
- Providing metadata⁶ either embedded within the record itself or in an accompanying digital file.

³ Government departments may also consult the *Procedures for the Transfer of Unpublished Information Resources of Enduring Value from Government of Canada Institutions to Library and Archives Canada* (2013).

⁴ This is a requirement for publications submitted to LAC on Legal Deposit, in accordance with the [Legal Deposit of Publications Regulations, section 2 \(a\)](#). For all other IREV, this applies only if the donor has the legal right to do so.

⁵ If the donor has the legal right to do so.

⁶ This is a requirement for publications submitted to LAC on Legal Deposit, in accordance with the [Legal Deposit of Publications Regulations, section 2 \(b\)](#). Generally, the preferred format of the metadata is a structured format (e.g. XML, CSV, DBF) to facilitate reuse. Furthermore, certain metadata standards may also be necessary such as ONIX 3.0 or Dublin Core for bibliographic metadata. Contact LAC to discuss metadata requirements prior to transfer.

9.1 Text Formats

Preferred Formats	Format Specifications
American Standard Code for Information Interchange Text (ASCII Text)	ISO/IEC 646:1991, <i>Information technology – ISO 7-bit coded character set for information interchange</i> : http://www.iso.org/iso/catalogue_detail.htm?csnumber=4777
Electronic Publication (EPUB3)	<i>International Digital Publishing Forum EPUB Version 3</i> : http://idpf.org/epub/30
Open Document Text Format (ODF)	ISO/IEC 26300:2006, <i>Information technology – Open Document Format for Office Applications (OpenDocument) v1.0</i> : http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=43485
Portable Document Format/Archival (PDF/A-1)	ISO 19005-1:2005, <i>Document management – Electronic document file format for long-term preservation – Part 1: Use of PDF 1.4 (PDF/A-1)</i> : http://www.iso.org/iso/catalogue_detail?csnumber=38920
Portable Document Format/Archival (PDF/A-2)	ISO 19005-2:2011, <i>Document management – Electronic document file format for long-term preservation – Part 2: Use of ISO 32000-1 (PDF/A-2)</i> : http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=50655
Unicode Text	RFC 3629: <i>UTF-8, A Transformation Format of ISO 10646</i> : http://tools.ietf.org/html/rfc3629 RFC 2781: <i>UTF-16: An Encoding of ISO 10646</i> : http://www.ietf.org/rfc/rfc2781.txt
Acceptable Formats	Format Specifications
Electronic Publication (EPUB2.0.1)	<i>International Digital Publishing Forum EPUB Version 2.0.1</i> : http://idpf.org/epub/201
Microsoft Word 97 Binary Document Format (doc)	[MS-XLS]: <i>Excel Binary File Format (.xls) Structure</i> : http://msdn.microsoft.com/en-us/library/cc313154(v=office.12).aspx
Microsoft Word Office Open XML (docx)	[MS-OI29500]: <i>Office Implementation Information for ISO/IEC 29500 Standards Support</i> : http://msdn.microsoft.com/en-us/library/ee908652%28v=office.12%29
Portable Document Format (PDF)	ISO 32000-1:2008, <i>Document management - Portable document format - Part 1: PDF 1.7</i> : http://www.iso.org/iso/catalogue_detail.htm?csnumber=51502

9.2 Presentation Formats

Preferred Formats	Format Specifications
OpenDocument Presentation Format (odp)	ISO/IEC 26300:2006, <i>Information technology - Open Document Format for Office Applications (OpenDocument) v1.0</i> : http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=43485
Portable Document Format Archival (PDF/A-1)	ISO 19005-1:2005, <i>Document management – Electronic document file format for long-term preservation – Part 1: Use of PDF 1.4 (PDF/A-1)</i> : http://www.iso.org/iso/catalogue_detail?csnumber=38920
Acceptable Formats	Format Specifications
Microsoft PowerPoint 1997-2007 Binary Format (ppt)	[MS-PPT]: <i>PowerPoint (.ppt) Binary File Format (.xls) Structure</i> : http://msdn.microsoft.com/en-us/library/cc313106(v=office.12).aspx
Microsoft PowerPoint Office Open XML Format (pptx)	[MS-OI29500]: <i>Office Implementation Information for ISO/IEC 29500 Standards Support</i> : http://msdn.microsoft.com/en-us/library/ee908652%28v=office.12%29

9.3 Email Formats⁷

Preferred Formats	Format Specifications
Internet Message Format (EML)	Internet Message Format: http://www.ietf.org/rfc/rfc2822.txt Multipurpose Internet Mail Extensions (MIME): http://tools.ietf.org/html/rfc2045 http://tools.ietf.org/html/rfc2046 http://tools.ietf.org/html/rfc2047 http://tools.ietf.org/html/rfc4288 http://tools.ietf.org/html/rfc4289 http://tools.ietf.org/html/rfc2049
MBOX Email Format	MBOX Email Format: https://tools.ietf.org/html/rfc4155 Multipurpose Internet Mail Extensions (MIME): http://tools.ietf.org/html/rfc2045 http://tools.ietf.org/html/rfc2046 http://tools.ietf.org/html/rfc2047 http://tools.ietf.org/html/rfc4288 http://tools.ietf.org/html/rfc4289 http://tools.ietf.org/html/rfc2049
Acceptable Formats	Format Specifications
Microsoft Outlook Item Message Format (MSG)	<i>[MS-OXMSG] Microsoft Outlook Item (.msg) File Format:</i> http://msdn.microsoft.com/en-us/library/cc463912(v=exchq.80).aspx
Microsoft Personal Folders Format (PST)	<i>[MS-PST]: Outlook Personal Folders (.pst) File Format:</i> http://msdn.microsoft.com/en-us/library/ff385210%28v=office.12%29.aspx

⁷ Email attachments are considered a component of the email and therefore the attachment does not have to meet the transfer standards specified by the format category that the attachment alone would fall under.

9.4 Formats for Still Images

This content category contains two subcategories: digital photographs and scanned text.

9.4.1 Digital Photographs

Preferred Formats	Format Specifications
Tagged Image File Format (TIFF), lossless	<i>TIFF Revision 6.0 Final — June 3, 1992, Adobe Systems Incorporated:</i> http://partners.adobe.com/public/developer/en/tiff/TIFF6.pdf
JPEG 2000 (JP2), lossless	ISO/IEC 15444-1:2004, <i>Information technology – JPEG 2000 image coding system: Core coding system:</i> http://www.iso.org/iso/catalogue_detail.htm?csnumber=37674
Portable Network Graphics (PNG)	ISO/IEC 15948:2004, <i>Information technology - Computer graphics and image processing -- Portable Network Graphics (PNG): Functional specification:</i> http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=29581
Acceptable Formats	Format Specifications
JPEG File Interchange Format (JFIF) with Joint Photographic Experts Group (JPEG) compression	ISO/IEC 10918-5:2013, <i>Information technology – Digital compression and coding of continuous-tone still images: JPEG Interchange File Format:</i> http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=54989 ISO/IEC 10918-1:1994, <i>Information technology – Digital compression and coding of continuous-tone still images: Requirements and guidelines:</i> http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=18902
Digital Imaging and Communications in Medicine (DICOM)	ISO standard 12052:2006, <i>Health informatics - Digital imaging and communication in medicine (DICOM) including workflow and data management:</i> http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=43218
Digital Negative (DNG), with preview JPEG image included	<i>Adobe Digital Negative (DNG) Specification Version 1.4.0.0:</i> http://www.images.adobe.com/www.adobe.com/content/dam/Adobe/en/products/photoshop/pdfs/dng_spec_1.4.0.0.pdf
Graphics Interchange Format (GIF)	<i>Graphics Interchange Format (sm) Version 89a:</i> http://www.w3.org/Graphics/GIF/spec-gif89a.txt

9.4.2 Scanned Text

Preferred Formats	Format Specifications
JPEG 2000 (JP2), lossless	ISO/IEC 15444-1:2004, <i>Information technology – JPEG 2000 image coding system: Core coding system</i> : http://www.iso.org/iso/catalogue_detail.htm?csnumber=37674
Portable Document Format/Archival (PDF/A), lossless	ISO 19005-1:2005, <i>Electronic document file format for long-term preservation – Part 1: Use of PDF 1.4 (PDF/A-1)</i> : http://www.iso.org/iso/catalogue_detail?csnumber=38920
Tagged Image File Format (TIFF), lossless	<i>TIFF Revision 6.0 Final – June 3, 1992 Adobe Systems Incorporated</i> : http://partners.adobe.com/public/developer/en/tiff/TIFF6.pdf
Acceptable Formats	Format Specifications
JPEG File Interchange Format (JFIF) with Joint Photographic Experts Group (JPEG) compression	ISO/IEC 10918-5: 2013, <i>Information technology – Digital compression and coding of continuous-tone still images: JPEG File Interchange Format (JFIF)</i> : http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=54989 ISO/IEC 10918-1:1994 <i>Information technology – Digital compression and coding of continuous-tone still images: Requirements and guidelines</i> : http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=18902
Plain text in combination with one of the above image formats	ISO/IEC 8859-1: 1988, <i>8-bit single-byte coded graphic character sets – Part 1: Latin alphabet No. 1</i> : http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=28245

9.5 Digital Audio Formats

Preferred Formats	Acceptable Codecs	Format Specifications
Broadcast Wave (BWF)	Linear Pulse Code Modulated Audio (LPCM)	<p>European Broadcast Union (EBU). Technical Specification of the Broadcast Wave Format (BWF) – Version 1: http://web.archive.org/web/20091229093941/http://tech.ebu.ch/docs/tech/tech3285.pdf</p> <p>Specification of the Broadcast Wave Format (BWF) - Version 2.0: https://tech.ebu.ch/docs/tech/tech3285.pdf</p>
Acceptable Formats	Acceptable Codecs	Format Specifications
Audio Interchange Format (AIFF)	Linear Pulse Code Modulated Audio (LPCM)	<p>Audio Interchange File Format: "AIFF" A Standard for Sampled Sound Files Version 1.3: http://www-mmsp.ece.mcgill.ca/Documents/AudioFormats/AIFF/Docs/AIFF-1.3.pdf</p>
Moving Pictures Expert Group (MPEG) MPEG-1 Layer 3, MPEG-2 Layer-3 (MP3)	MP3enc, Lame	<p>ISO/IEC-11172-3: 1993, Information technology – Coding of moving pictures and associated audio for digital storage media at up to about 1,5 Mbit/s – Part 3: Audio: http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=22412</p> <p>ISO/IEC 13818-3:1998 Information technology – Generic coding of moving pictures and associate audio information – Part 3: Audio: http://www.iso.org/iso/home/store/catalogue_ics/catalogue_detail_ics.htm?csnumber=26797</p>
MPEG-4 AAC Advanced Audio Coding (AAC)	n/a	<p>ISO/IEC 14496-3:2009, Information technology - Coding of audio-visual objects - Part 3: Audio: http://www.iso.org/iso/iso_catalogue/catalogue_ics/catalogue_detail_ics.htm?csnumber=53943</p>
WAVEform Audio (WAV)	Linear Pulse Code Modulated Audio (LPCM)	<p>Multimedia Programming Interface and Data Specifications 1.0: http://www-mmsp.ece.mcgill.ca/Documents/AudioFormats/WAVE/Docs/riffmci.pdf</p>

9.6 Formats for Digital Moving Images

This content category contains two subcategories: digital cinema and digital video.

9.6.1 Digital Cinema

Preferred Formats	Acceptable Codecs	Format Specifications
Digital Cinema Distribution Master (DCDM)	n/a	Digital Cinema Initiatives, <i>DCI Specification Version 1.2, 2012</i> : http://www.dcinovies.com/specification/index.html SMPTE 428-1-2006: <i>D-Cinema Distribution Master (DCDM) – Image Characteristics</i> : http://standards.smpte.org/
Digital Moving Picture Exchange Bitmap (DPX)	Uncompressed	SMPTE ST 268:2003, <i>File Format for Digital Moving-Picture Exchange (DPX) Version 2.0</i> : http://standards.smpte.org/
Acceptable Formats	Acceptable Codecs	Format Specifications
Digital Cinema Package (DCP) Unencrypted Interop or SMPTE compliant	JPEG 2000 (as outlined by the DCI specifications)	Digital Cinema Initiatives, <i>DCI Specification Version 1.2, 2012</i> : http://www.dcinovies.com/specification/index.html

9.6.2 Digital Video

Preferred Formats	Acceptable Codecs	Format Specifications
Audio Video Interleaved Format (AVI)	Uncompressed 4:2:2	AVI RIFF File Reference: http://msdn.microsoft.com/en-us/library/ms779636.aspx
Material Exchange Format (MXF) OP1a	JPEG 2000 lossless compression	SMPTE ST 377-1:2011, <i>Material Exchange Format (MXF) File Format Specification</i> : http://standards.smpete.org/ ISO/IEC 15444-1:2004, <i>Information technology – JPEG 2000 image coding system: Core coding system</i> : http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=37674
Quicktime (MOV)	Uncompressed 4:2:2	QuickTime File Format Specification: https://developer.apple.com/library/mac/#documentation/QuickTime/QTFF/QTFFPreface/qtffPreface.html
Acceptable Formats	Acceptable Codecs	Format Specifications
Audio Video Interleaved Format (AVI)	JPEG 2000 DV-NTSC AVC/H.264	AVI RIFF File reference: http://msdn.microsoft.com/en-us/library/ms779636.aspx ISO/IEC 15444-1:2004, <i>Information technology – JPEG 2000 image coding system: Core coding system</i> : http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=37674 <i>Microsoft NTSC DV-AVI File reference</i> : http://msdn.microsoft.com/en-us/library/windows/desktop/dd407250%28v=vs.85%29.aspx H.264: <i>Advanced video coding for generic audiovisual services</i> : http://www.itu.int/rec/T-REC-H.264-201304-I/en
MPEG-2 Video (MPEG2)	n/a	ISO/IEC 13818-2:2013, <i>Information technology - Generic coding of moving pictures and associated audio information Part 2: Video</i> : http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=61152

MPEG 4	AVC/H.264	<p>ISO/IEC 14496-14:2003, <i>Information technology – Coding of audio-visual objects – Part 14: MP4 file format</i>: http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=38538</p> <p>H.264: <i>Advanced video coding for generic audiovisual services</i>: http://www.itu.int/rec/T-REC-H.264-201304-I/en</p>
QuickTime File Format (MOV)	JPEG 2000 DV-NTSC AVC/H.264 Apple ProRes Avid DNxHD	<p><i>QuickTime File Format Specification</i>: https://developer.apple.com/library/mac/#documentation/QuickTime/QTFF/QTFFPreface/qtffPreface.html</p> <p>ISO/IEC 15444-1:2004, <i>JPEG 2000 image coding system</i>: http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=37674</p> <p>H.264: <i>Advanced video coding for generic audiovisual services</i>: http://www.itu.int/rec/T-REC-H.264-201304-I/en</p> <p><i>Apple ProRes White Paper October 2012</i>: http://images.apple.com/finalcutpro/docs/Apple_ProRes_White_Paper_October_2012.pdf</p> <p><i>Avid DNxHD/SMPTE VC-3</i>: http://standards.smpte.org/</p>
Windows Media Video 9 File Format (WMV)	VC-1	<p><i>Advanced Systems Format (ASF) Specification</i>: http://msdn.microsoft.com/en-us/library/bb643323.aspx</p> <p>SMPTE ST 421:2013, <i>VC-1: Compressed Video Bitstream Format and Decoding Process</i>: http://standards.smpte.org/content/978-1-61482-770-2/st-421-2013/SEC1.abstract?sid=1594d1cc-1b4e-47c6-a2be-006e86cd49f1</p>

9.7 Geospatial Formats

Preferred Formats	Format Specifications
Band Interleaved by Line (BIL)	<i>BIL, BIP, and BSQ raster files:</i> http://webhelp.esri.com/arcgisdesktop/9.3/index.cfm?id=2915&pid=2910&topicname=BIL, BIP, and BSQ raster files
Band Interleaved by Pixel	<i>BIL, BIP, and BSQ raster files:</i> http://webhelp.esri.com/arcgisdesktop/9.3/index.cfm?id=2915&pid=2910&topicname=BIL, BIP, and BSQ raster files
Band Interleaved Sequential (BSQ)	<i>BIL, BIP, and BSQ raster files:</i> http://webhelp.esri.com/arcgisdesktop/9.3/index.cfm?id=2915&pid=2910&topicname=BIL, BIP, and BSQ raster files
Digital Elevation Model (DEM)	<i>USGS, Part 1: General and Part 2: Specifications, Standards for Digital Elevation Model:</i> http://nationalmap.gov/standards/demstds.html
Environmental Systems Research Institute (ESRI) Arc/Info ASCII Grid	<i>ESRI ASCII Raster Format:</i> http://help.arcgis.com/en/arcgisdesktop/10.0/help/index.html#/ESRI_ASCII_raster_format/09t0000000z000000/ http://webhelp.esri.com/arcgisdesktop/9.1/index.cfm?id=886&pid=885&topicname=ASCII%20to%20Raster%20(Conversion) http://resources.esri.com/help/9.3/arcgisengine/java/GP_ToolRef/spatial_analyst_tools/esri_ascii_raster_format.htm
Environmental Systems Research Institute (ESRI) Shapefile (SHP)	<i>ESRI Shapefile Technical Description:</i> http://www.esri.com/library/whitepapers/pdfs/shapefile.pdf
GeoTiff	<i>GeoTiff Format Specification, Version 1.8.2, Revision 1.0, 2000:</i> http://www.remotesensing.org/geotiff/spec/geotiffhome.html
Geography Markup Language (GML)	ISO 19136:2007 & Version 3.2, <i>OpenGIS Geography Markup Language (GML) Encoding Standard 07-036:</i> http://www.opengeospatial.org/standards/is
Keyhole Markup Language (KML)	<i>Open Geospatial Consortium Inc. OGC KML 07-147r2:</i> http://www.opengeospatial.org/standards/kml/

Acceptable Formats	Format Specifications
Canadian Council on Geomatics Interchange Format (CCOGIF)	Canadian Council on Geomatics, Standard File Exchange Format For Digital Spatial Data, Version #2.3, October 1994
Digital Lines Graphics – Level 3 (DIG-3)	USGS, Part 1: General and Part 2: Specifications, Standards for Digital Line Graphs: http://nationalmap.gov/standards/dlgstds.html
Environmental Systems Research Institute (ESRI) Export Format (E00)	Reverse engineered specification, Arc/Info Export (E00) Format Analysis: http://avce00.maptools.org/docs/v7_e00_cover.html
Geospatial PDF	ISO 32000-1:2008, <i>Document management – Portable document format – Part 1: PDF 1.7</i> : http://www.iso.org/iso/catalogue_detail.htm?csnumber=51502
International Hydrographic Organization (IHO) S-57	<i>IHO Transfer Standard for Digital Hydrographic Data. Edition 3.1 - November 2000 Special Publication No. 57</i> : http://www.iho.int/iho_pubs/standard/S-57Ed3.1/31Main.pdf
TerraGo GeoPDF	Open Geospatial Consortium Inc. OGC 08-139r2: https://portal.opengeospatial.org/modules/admin/license_agreement.php?suppressHeaders=0&access_license_id=3&target=http://portal.opengeospatial.org/files/?artifact_id=33332

9.8 Computer Aided Design Formats

Preferred Formats	Format Specifications
AutoDesk's Drawing File	<p>Open version of the specification available via the Open Design Alliance: http://www.opendwg.org/</p>
AutoDesk's Drawing Interchange File Format/Data eXchange Format (DXF)	<p>AutoCAD DXF, v.u.28.1.01 http://usa.autodesk.com/adsk/servlet/item?siteID=123112&id=12272454&linkID=10809853</p>
Acceptable Formats	Format Specifications
Portable Document Format/ Engineering (PDF/E)	<p>ISO 24517-1:2008, <i>Document management - Engineering document format using PDF - Part 1: Use of PDF 1.6 (PDF/E-1)</i>: http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=42274</p>
Standard for the Exchange of Product Model Data (STEP)	<p>ISO 10303-21:2002, <i>Industrial automation systems and integration - Product data representation and exchange - Part 21: Implementation methods: Clear text encoding of the exchange structure</i>: http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=33713</p> <p>ISO 10303-28:2007, <i>Industrial automation systems and integration -- Product data representation and exchange -- Part 28: Implementation methods: XML representations of EXPRESS schemas and data, using XML schemas</i>: http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=40646</p>

9.9 Formats for Data Sets

Tabular data from databases and spreadsheets must meet the following requirements:

- Each record must contain an end-of-record marker;
- Each field within a file must be defined with the same fixed width;
- Each record must be defined with the same logical record length;
- All fields within a record in a database, or tuples in a relational database, should have the same logical format;
- A record should not contain nested repeating groups of data;
- Every file must be accompanied by documentation that specifies the field names and the field definitions.⁸

Preferred Formats	Format Specifications
American Standard Code for Information Interchange Text (ASCII Text)	ISO/IEC 646:1991, <i>Information technology - ISO 7-bit coded character set for information interchange</i> : http://www.iso.org/iso/catalogue_detail.htm?csnumber=4777
Comma Separated Value (CSV)	<i>Common Format and MIME Type for Comma-Separated Values (CSV) Files</i> : http://tools.ietf.org/html/rfc4180
Acceptable Formats	Format Specifications
dBASE Table File Format (DBF)	<i>Data File Header Structure for the dBASE Version 7 Table Files</i> : http://www.dbase.com/Knowledgebase/INT/db7_file_fmt.htm
Extended Binary Coded Decimal Interchange Code (EBCDIC)	<i>IBM EBCDIC Code Page 0037</i> : http://www-01.ibm.com/software/globalization/cp/cp00037.html
Microsoft Excel Office Open XML	[MS-OI29500]: <i>Office Implementation Information for ISO/IEC 29500 Standards Support</i> : http://msdn.microsoft.com/en-us/library/ee908652%28v=office.12%29
Microsoft Excel 97 Binary Document Format (xls)	[MS-XLS]: <i>Excel Binary File Format (.xls) Structure</i> : http://msdn.microsoft.com/en-us/library/cc313154(v=office.12).aspx
OpenDocument Format Spreadsheet (ODS)	ISO/IEC 26300:2006, <i>Information technology - Open Document Format for Office Applications (OpenDocument) v1.0</i> : http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=43485

⁸ Please clarify the specific documentation requirements for data sets with the LAC representative responsible for the transfer.

10. Roles and responsibilities

Responsibility for administering these Guidelines rests with the Directors General of the relevant functional areas.

Directors are responsible for implementing these Guidelines within their management areas.

LAC staff involved in the acquisition, stewardship and reappraisal of digital IREV are responsible for communicating and operationalizing these Guidelines.

Donors are to adhere to these Guidelines and consult with LAC on any matters that may impede their ability to comply with these Guidelines.

11. Monitoring, evaluation and review

The functional area responsible for acquisition will monitor application of these Guidelines and report on compliance.

Evaluation and review of these Guidelines will be undertaken every 3 years by representatives of the branches responsible for acquisition and stewardship, or earlier if requested by senior management.

12. Consequences

Non-compliance with these Guidelines will have a negative impact on acquisition, stewardship and reappraisal activities and results.

Consequences for non-compliance with these Guidelines may include initial or full rejection of proposed file transfers, or corrective measures, at the discretion of LAC staff responsible for the acquisition of IREV. Corrective measures may include any actions deemed appropriate and acceptable under the circumstances.

13. Information

Please address any questions about these Guidelines to:

Director General
Evaluation and Acquisition Branch
Library and Archives Canada
550 de la Cité Boulevard
Gatineau, Québec
K1A 0N4

Appendix A: Definitions

Acceptable format: a file format that meets LAC's minimum requirements for sustainability. This format may require LAC to perform some preservation actions on ingest to ensure their long-term sustainability.

Bitmap: an image created from a series of bits and bytes that form pixels. Each pixel carries a value that defines a bits/bytes colour or greyscale. Such images are also known as raster images.

Codec: hardware or software capable of encoding and/or decoding a data stream for transmission. When used with digital audio or video, the term codec refers to the digital signal encapsulated in a wrapper.

Container format: a format that can contain and support various types or layers of audio, video, still imagery and their associated metadata. For the data stream to be properly interpreted, the information must be encapsulated or wrapped together. The wrapper refers to a particular way of storing and synchronizing data content into a single file.

Compression: the encoding of information using fewer bits than in the original. There are two forms of data compression – lossless and lossy. A lossless compression technique discards no information. It looks for more efficient ways to represent data, while making no compromises in accuracy. Lossy compression accepts some degradation in the data to achieve smaller file sizes. Because of this degradation in quality, lossy compression should be avoided.

Computer Aided Design (CAD): vector programs used to create animations that represent two- and three-dimensional surfaces of inanimate objects. CAD and vector graphics programs can output binary and XML formats.

Data sets: data stored in defined fields such as databases and spreadsheets.

Database formats: organized collections of data that conform to a logical structure. Database formats are determined by data models that describe specific data structures used to model an application and generally include navigational, relational, and hybrid models.

Digital audio: file formats that encode recorded sound as machine readable files by converting acoustic sound waves into digital signals. Digital audio formats are generally composed of both a wrapper format and an encoding

method or codec. Audio file stream encodings are independent of the audio container file format.

Digital cinema: both born-digital cinematic productions and digital moving image files created by digitizing motion picture film.

Digital moving images: a sequence of bitmap digital images displayed in rapid succession at a constant rate, giving the appearance of movement. Digital moving image file formats function as containers or wrappers to provide storage areas for any moving image essence, associated audio essence (if present), as well as metadata. Moving image essence data contained within a given wrapper file format is encoded for playback using a specific codec. The parameters of the codec employed determines the presence and method of compression that was used to store the digital moving image data within the wrapper. This category includes two subcategories: digital cinema and digital video.

Digital photographs: both still photographs produced by digital cameras as well as scanned images of photographic prints, slides, and negatives.

Digital rights management technologies: technologies to prevent unauthorized use or reproduction of digital content and devices.

Digital video: both born-digital video and digital files created by digitizing video from an analog source.

Email: electronic communication transmitted over the Simple Mail Transfer Protocol (SMTP) between two or more accounts. Email is composed of a header, message body and attachments. The header is structured metadata that establishes the provenance of the record. Data that must be present is: sender name and address; names and addresses of all recipients; sent date; and, received date. The message body is the intellectual content of the message. Attachments are any additional objects sent with the email.

Encapsulating format: see container format.

Encryption: the use of an algorithm to render a file unreadable. A decryption key is required to undo the work of the algorithm.

Enduring value: the quality of having continuing archival or historical usefulness or significance to Canadian society.

End-of-record marker: in a file varies in accordance with the operating system this is used to create the file. In a MAC OS environment a carriage return (CR - ASCII code 0x0D) is placed at the end of a record. In a DOS or Windows OS environment a CR+ a Line Feed (LF – ASCII code 0x0A) is placed at the end. In UNIX only a LF appears at the end.

File format: specific pattern or structure that organizes and defines data. Some formats contain only one stream of uncompressed data, others may contain codecs to encode and compress the data, and others may support several streams of media.

Geospatial: data may be contained within a database to enable analysis across the datasets (e.g. geo-database), united within a complex file format structure where one geospatial file is comprised of several distinct, but related, formats (e.g. shapefile), or contained within a single file (e.g. GML).

Information resources: any documentary material, published or unpublished, regardless of communications source, information format, production mode or recording medium.

Information resources of enduring value (IREV): information resources that have long-term importance and relevance to Canadian society.

Metadata: data about other data.

Migration: the movement of digital information from one software/hardware environment/storage medium to another as standards and technology evolve.

Preferred format: a file format that is readily usable and has been identified by LAC as possessing a high degree of long-term sustainability. This format requires little or no immediate management to achieve appropriate levels of preservation.

Presentation format: a format that conveys graphical information to audiences as a slide show.

Raster image: see bitmap.

Scanned text: a photograph of a printed page produced by either a digital camera or scanner.

Spreadsheets: tables made up of columns and rows that contain cells of data. Relationships between cells can be pre-defined as mathematical formulas.

Still images: files that are sampled and bitmapped as a grid of rectangular dots, picture elements (pixels) or points of color.

Stewardship: the responsible management of IREV in one's care, custody, control or ownership so that it can be passed on to future generations.

Sustainability: ensuring that the documentary heritage acquired and managed by LAC is accessible over time, including giving consideration to its one-time or ongoing stewardship requirements and to LAC's resource capacity. In the context of these Guidelines sustainability is tied to the suitability of a format to preserve encoded information over time. Factors that contribute to a format's sustainability include quality, stability, potential longevity and industry acceptance.

Text: there are two general types of text: plain and formatted. Formatted text files contain encoded ASCII data and format definitions that display the information in a defined pattern. Plain text files contain encoded ASCII or Unicode data that has no formatting or layout code to influence the presentation of the data.

Unacceptable format: a format that does not meet the minimum requirements to be considered sustainable by LAC.

Vector graphics: digital images made up of object-oriented images that use the geometry of points, lines, curves and polygons to represent images.

Wrapper: see container format.

Appendix B: Bibliography

Library of Congress. *Sustainability of Digital Formats*.

<http://www.digitalpreservation.gov/formats/>. Accessed August 20, 2013.

National Archives and Records Administration. *NARA Bulletin 2013-XX Revised Format Guidance for the Transfer of Permanent Electronic Records*. 2013.

National Archives (UK). *Suitable file formats for transfer of digital records to The National Archives*. September 2011.

<http://www.nationalarchives.gov.uk/documents/information-management/file-formats-for-transfer.pdf>. Accessed August 20, 2013.