INITIAL DRAFT

## Library and Archives Canada

# Guidelines for Computer File Types, Interchange Formats and Information Standards

**Electronic Records Development Division**
**Government Records Branch**
**Library and Archives Canada**

Version 0.6
No-Date, 2003

Document Identification

| | |
|---|---|
| Title | Library and Archives Canada: Guidelines on Computer File Types, Interchange Formats and Information Standards |
| Author | David L. Brown |
| Subject | Electronic Record File Formats and Interchange Formats |
| Description | Suggested formats for creating and transferring electronic records to Library and Archives Canada |
| Publisher | Library and Archives Canada |
| Contributor | Mike Swan |
| Date | No-Date, 2003 |
| Type | Text |
| Format | Microsoft Word 2000 |
| Identifier | Version 0.6 |
| Source | |
| Language | English |
| Relation | |
| Coverage | |
| Rights | Intellectual property rights – owned by Canada<br>© Copyright – Her Majesty the Queen in Right of Canada - 2003 |

Standard Document Identification – Dublin Core Metadata Element Set Version 1.1 1999-07-02

Document Change Control

| Revision Number | Date of Issue | Author(s) | Brief Description of Change |
|---|---|---|---|
| Version 0.1 | 13 June, 2003 | Mike Swan | Original |
| Version 0.2 | 7 July, 2003 | David Brown | Review and inclusion of Geomatics |
| Version 0.3 | 7 August, 2003 | Mike Swan, David Brown | Inclusion of other formats and deletion of specifications. |
| | 25 August, 2003 | David Brown | Modification of Still Imagery Section |
| Version 0.4 | 25 September, 2003 | David Brown | Major reworking of the Introductory Section, and Inclusion of Presentation/Character Set Section. |
| Version 0.5 | 17 October 2003 | David Brown | Major modification of entire document based on comments from within GRB and select people from GPC. |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |

# Table of Contents

# 1 Introduction

## 1.1 Purpose and Scope

This document identifies computer file types; interchange formats and information standards that the Library and Archives Canada (LAC) is recommending to facilitate the interoperability of digital information in the Government of Canada (GoC). This document focuses upon specific facets related to information interoperability that enable the sharing and exchange of information between the LAC and other agencies in the GoC. The file types and interchange formats cited in this document are intended to cover a number of data and information types; including computer generated digital audio, digital still imagery, digital video, documents - textual, email, geospatial data, structured data - databases and spreadsheets, and technical computer aided design (CAD) drawings. The information standards address data presentation issues.

Although the LAC has the technological capability to handle the entire set of file formats and standards identified in this document, they have been categorized into those that are "recommended" for use and those that are "acceptable" for use. Those identified as "recommended" are being promoted by the LAC for the creation of computer-generated information from a purely technical rationale. Recommended file types and interchange formats are also those that are preferred by the LAC for the transfer of digital information to its control after its operational business value to an organization has ceased. These file types and interchange formats are also those the LAC is promoting for the exchange of digital information in the GoC. Computer file types, interchange formats and information standards that are identified as being "acceptable" are suitable only if certain criteria are met.

When GoC departments and agencies have archival information contained in computer files or interchange formats other than those specified in this document, they must consult the LAC to determine whether it is an acceptable format prior to transferring the information.

## 1.2 Background

The Treasury Board of Canada Secretariat (TBS) develops GoC information management (IM) policy and its implementation in the GoC is enhanced through guidance from the Library and Archives Canada. Under the auspices of the *National Archives of Canada Act*, the LAC has responsibility for preserving the collective memory of the Nation and the Government of Canada. Under Section four (4) of the *Act*, the Archives can acquire 'records' from the 'private and public' sectors that it considers to be of national significance. Under the definition of a record in the *Act* this includes 'machine readable record[s]'.

The preservation of digital information is an issue of enormous importance. The GoC is creating and storing terabytes of digital information, most of which is stored in a variety of logical record formats. The efficient operational management of these records is critical to ensure the

availability of the information to future generations of government policy and decision makers, and to conduct various types of government research.

The long-term access to data created by the GoC will be compromised unless policies, procedures and tools are created and implemented to ensure their effective management and eventual preservation. Electronic records are by their nature more fragile than paper records and permanent access to their content is more vulnerable to change or loss. Access to digital information is dependent upon software and hardware that can change rapidly over time. It is very common for software and hardware to become obsolete within a few years of their release. The preservation of digital bits is easily achieved, but if the computer platforms and software applications needed to interpret the information are no longer available, the 'value' this information represents will be lost forever.

Working in partnership with the library and archival communities, data producers in the GoC need to standardize and adopt organizational policies and practices to govern the creation, use, retention, dissemination, preservation, and disposition of digital information to ensure its authenticity and integrity for as long as laws, regulations or government policies and directives require it.

## 1.3 Concept

The LAC has created this document to provide guidance to departments and agencies in the GoC on computer file types, interchange formats and information standards that should be considered during the creation of digital information. The adoption of these formats and standards will facilitate information exchange between departments, provide a basis for the implementation of common IM practices throughout the GoC and ensure the preservation of 'records of value' for future generations of Canadians. This document is only intended to identify formats and information standards that are recommended or accepted by the LAC for the conduct of government business. Technical specifications for the application of specific formats and standards will be developed and released as appendices to this document as they are defined.

Standardizing the formats for the creation, use and transfer of digital information is an essential element of the long-term preservation process. A platform independent, industry supported standard logical format should allow reliable access to electronic records for a period of five years before the information must be migrated to a new format. The physical medium upon which the records are stored also plays a vital role in the preservation equation, but this issue will not be explicitly addressed in this document. Migration procedures are very costly to implement and could expose the information to the risks of degradation and loss. As a result, limiting the frequency of data migration and examining the associated risks should be a required component of any information management and preservation strategy.

In selecting the file types, interchange formats and information standards, the LAC attempted to balance the requirements for quality, stability, potential longevity and industry acceptance. Where possible, a preference was placed on the selection of non-proprietary national and

international interchange formats, information standards, or *De facto* standard industry formats and file types. *De facto* standard formats are widely used and recognized formats and file types 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 terms of application, publicly available specifications are being promoted for GoC use to eliminate any potential reliance on the fate of any specific company recommendation. The formats appear in alphabetical order within the relevant areas.

## 1.4 Updates

In order to maintain the currency of this document, the information presented herein will be reviewed and updated regularly to reflect the operational requirements that exist in the GoC and to meet the challenges of evolving technological advancements. People are invited to comment on the contents of new document versions as they are released. To direct comments, please see the Enquires section (1.5.4).

## 1.5 Guidance

This policy should be read in conjunction with relevant GoC legislation, policies and guidelines.

### 1.5.1 Legislation

*Access to Information Act*
*Canada Evidence Act*
*Copyright Act*
*Criminal Records Act*
*Emergency Preparedness Act*
*Financial Administration Act*
*National Archives of Canada Act*
*National Library Act*
*Official Languages Act*
*Official Secrets Act*
*Personal Information Protection and Electronic Documents Act*
*Privacy Act*
*Statistics Act*

### 1.5.2 Related Treasury Board of Canada Policies

*Common Look and Feel for the Internet: Standards and Guidelines*
*Common Services*
*Communications*
*Data Matching*
*Electronic Authorization and Authentication*
*Enhanced Management Framework*
*Evaluation*
*Government Security*

*Internal Audit*
*Management of Government Information*
*Management of Information Technology*
*Policy, Guidelines and Standards for Public Key Infrastructure Management*
*Policy on using the Official Languages on Electronic Networks and other official languages policies*
*Privacy and Data Protection*
*Privacy Impact Assessment*

### 1.5.3 Related Library and Archives Canada Policies

*Electronic Publishing: Guide to Best Practices for Canadian Publishers, Version 1.0*
*Guidelines for Managing Recorded Information in a Minister's Office*
*Guidelines for Records Created Under a Public Key Infrastructure Using Encryption and Digital Signatures*
*Managing Audio-visual Records of the Government of Canada*
*Managing Cartographic, Architectural and Engineering Records in the Government of Canada*
*Managing Documentary Art Records of the Government of Canada*
*Managing Electronic Records in an Electronic Work Environment*
*Managing Photographic Records in the Government of Canada*
*Managing Shared Directories and Files*
*Protecting Essential Records*
*Federal Records Centers User Guide*

### 1.5.4 Enquiries

Enquiries about the content of this document should be directed to:

Electronic Records Development Division
Government Records Branch
Library and Archives Canada
344 Wellington St.
Ottawa, ON, Canada
K1A 0N3
613-944-4644 (Voice)
613-947-1500 (FAX)
imgi@archives.ca

# 2 Presentation

## 2.1 Character Sets

### 2.1.1 Recommended

#### 2.1.1.1 American Standard Code for Information Interchange (ASCII) [ISO/IEC 8859-1:1998 (Latin-1)]

The LAC supports the use of the ISO/IEC 8859-1:1998 ASCII character set for encoding. The standard defines a set of 256 characters where each character is defined using 8-bit binary numbers.

*Version:* ISO/IEC 8859-1:1998

http://www.iso.org/iso/en/CatalogueDetailPage.CatalogueDetail?CSNUMBER=28245&IC1=35&ICS2=40&ICS3

#### 2.1.1.2 Extended Binary Coded Decimal Interchange Code (EBCDIC)

EBCDIC is an encoding schema that is used by IBM mainframe computers. The character set was developed in the 1960s and similar to ASCII, it uses an 8 bit binary code to represent up to 256 characters. The character set comes in six slightly different forms, but it is still being used today on IBM mainframes. Detailed information on EBCDIC can be found in the IBM publication *IBM Character Data Representation Architecture, Reference and Registry, SC09-2190-00,* December 1996.

#### 2.1.1.3 Unicode Version 3.0 UTF-8 [ISO/IEC 10646-1:2000]

The LAC supports the Unicode version 3.0 standard that defines a multi-octet character set called the Universal Character Set (UCS). Unicode 3.0 UTF-8 (UCS Transformation Format - 8) provides a unique number for up to 49,194 characters, regardless of the platform, program or language. Unicode 3.0 has been updated by later versions of the standard. These updates do not replace the bulk of the existing material of Unicode 3.0. These revisions add characters, correct or extend the character properties in the Unicode Character Database or have significance for the interpretation of some aspects of the standard. The Unicode standard is recommended by the LAC because it provides the default UCS encoding scheme for HTML, SGML, XHTML and XML.

*Versions:* 1.0, 1.1, 2.0, 2.1, 3.0, 3.1, 3.2, and 4.0

http://www.unicode.org/book/u2.html

# 3 File Types and Interchange Formats

## 3.1 Digital Audio

### 3.1.1 Recommended

#### 3.1.1.1 Audio Interchange File Format (AIFF)

Audio IFF provides a standard for storing sampled sounds. The format is quite flexible, allowing for the storage of mono or multi-channel sampled sounds at a variety of sample rates and sample widths. It is primarily an interchange format and is intended for use with a large variety of computers, sampled sound instruments, sound software applications, and high fidelity recording devices. It does not support data compression, so AIFF files are often very large. Audio IFF is widely used in professional programs that process digital audio waveforms.

*Versions*: 1.1, 1.2 and 1.3

http://preserve.harvard.edu/standards/Audio%20IFF%20Specification%201%203.pdf

#### 3.1.1.2 Windows Audio Visual (WAV)

Microsoft and IBM developed the WAV format jointly. Support for WAV files was built into Windows 95, making it the *De facto* standard for sound on PCs. It is now supported by nearly all Windows applications. The format supports many bit resolutions, sample rates, audio channels and a number of lossless compression methods. WAV is widely used in professional programs that process digital audio waveforms. As a long-standing digital audio format, WAV remains the *De facto* standard for audio files in use today. As such, the LAC supports and recommends the use of the WAV file format.

### 3.1.2 Acceptable

#### 3.1.2.1 MPEG –1: Layer 3 (MP3)

The MP3 format is a compression system for music that reduces songs by a factor of 10 to 14 without changing the quality of a song's sound. The compression method used is lossy, thus data from the original file will be lost during compression. The standard has been widely adopted by both software manufactures and users. The MP3 standard is available at: http://mpeg.telecomitalialab.com/standards/mpeg-1/mpeg-1.htm

#### 3.1.2.2 Musical Instrument Digital Interface (MIDI)

MIDI is a standard adopted by the electronic music industry for controlling devices such as synthesizers and sound cards that emit music. At a minimum, a MIDI representation of a sound includes the note's pitch, length and volume, but it also can include other characteristics like attack and delay time. MIDI is a *De facto* standard for communication between musical instruments and the source of music for PC games. The MIDI specification is available from:

http://www.midi.org/about-midi/specinfo.shtml

#### 3.1.2.3 Real Audio (RM/RA)

RealAudio was the first streaming media product for the Internet and has become a *De facto* standard for network audio. It uses a lossy compression format that first deletes the very high and very low frequencies that cannot be detected by the human ear. It then removes as much data as possible, while keeping certain frequencies intact. More information about Real Audio can be found at:
http://www.realnetworks.com/resources/howto/audio_video/audio.html

## 3.2 Digital Still Imagery

### 3.2.1 Recommended

#### 3.2.1.1 International Telecommunication Union-Telecommunication Standardization Sector (ITU-T) T.4 and T.6

Originally known as Comité Consultatif International Téléphonique et Télégraphique (CCITT) Group 3 and Group 4, the ITU-T recommendations T.4 and T.6 are compression methods that were developed for the lossless compression of imagery data. Loseless refers to compression techniques where no data are lost during the data compaction process. The LAC prefers that digital images remain uncompressed. When it is impractical to store or transfer uncompressed files, the LAC recommends the use of a lossless compression method. The developers of fax machines originally adopted CCITT compression techniques, but the makers of general document storage and retrieval systems now use them heavily. The compression method takes advantage of an image's tendency to consist of a small number of black pixels on a white background. The encoding method involves changing the runs of white and black pixels into code words that are stored in a Huffman table. A Huffman table is essentially a codebook that allows one to decode a body of data.
*Versions:* T.4
http://www.itu.int/rec/recommendation.asp?type=items&lang=e&parent=T-REC-T.4-199904-I
T.6
http://www.itu.int/rec/recommendation.asp?type=items&lang=e&parent=T-REC-T.6-198811-I

#### 3.2.1.2 Portable Network Graphics (PNG)

PNG is an extensible file format for the lossless, compressed, portable storage of raster image data. Raster images are based on grids of dots, or pixels, where each pixel is represented by a numeric colour code. The format was designed to provide a patent-free, high quality replacement for the GIF file format (see below). PNG supports the indexed-colour, grayscale, and true-colour image modes, as well as an optional alpha channel. More information on PNG can be found at http://www.libpng.org/pub/png/.
*Versions*: 1.0 http://www.libpng.org/pub/png/spec/1.0/
1.1 http://www.libpng.org/pub/png/spec/1.1/
1.2 http://www.libpng.org/pub/png/spec/

#### 3.2.1.3 Tagged Image File Format (TIFF)

TIFF is the LAC's preferred standard for describing and storing raster image data from scanners, faxes and digital photography applications. It is capable of describing bilevel, grayscale, palette-colour, and full-colour images in several colour spaces. TIFF is extensible, portable and does not favour a particular computer operating system, compiler or processor. The TIFF copyright is owned by Adobe, but the specification is openly available and is supported by most conversion tools and photography software applications.

*Versions*: Revision 6.0 http://partners.adobe.com/asn/developer/pdfs/tn/TIFF6.pdf
Revision 5.0 http://palimpsest.stanford.edu/bytopic/imaging/std/tiff5.html

### 3.2.2 Acceptable

#### 3.2.2.1 Graphics Interchange Format (GIF)

CompuServe released GIF in 1987 as a free and open specification for the storage of raster imagery and to facilitate the exchange of digital imagery between different computer platforms and operating systems. Since 1987, the GIF format has become one of the most widely used formats for the storage of imagery data. GIF images are compressed using the Unisys patented Lempel Ziv Welch (LZW) compression and decompression technology to reduce file sizes. The Canadian LZW patent expires on July 7, 2004. The patent does not cover GIF files and there is no risk associated with the distribution and storage of GIF files by the GoC. The patent impacts only software using the LZW compression algorithm.

*Versions:* 87a http://www.w3.org/Graphics/GIF/spec-gif87.txt
89a http://www.whisqu.se/per/docs/graphics54.htm

#### 3.2.2.2 Joint Photographic Experts Group (JPEG) [ISO/IEC 10918-1:1994]

JPEG is a standardized lossy image compression mechanism that is designed for compressing full-colour and grayscale images. The International Organization for Standardizaton (ISO) standardized the JPEG compression format in 1990. The format is designed upon a specification that controls the compression ratio of the associated output image. It uses lossy compression that is designed to exploit the fact that humans perceive small colour changes less accurately than small changes in brightness. JPEG works well for photographs and artwork, but does not accurately represent lettering, cartoons or line drawings. In addition, ISO has developed a new version of JPEG know as JPEG2000. This standard was released in January of 2001, but it is still not widely used. As a result, the LAC is monitoring developments with respect to the use of JPEG2000. For more information on JPEG, refer to the ISO/IEC 10918-1:1994 standard or the JPEG FAQ:
http://www.faqs.org/faqs/jpeg-faq/
For further information on JPEG2000, refer to:
http://www.jpeg.org/JPEG2000.html

#### 3.2.2.3 JPEG File Interchange Format (JFIF)

JFIF is a simplified format that enables JPEG compressed images to be exchanged between a wide variety of computer platforms and software applications. JPEG JFIF is a file format that was created by the Independent JPEG Group (IJG) for the single transport of JPEG compressed images. When most people refer to JPEG, JFIF is the file format to which they are referring. JFIF is fully compliant with the JPEG standard.

*Versions:* 1.02 (Sept., 1992) http://www.w3.org/Graphics/JPEG/jfif3.pdf

## 3.3 Digital Video

### 3.3.1 Recommended

#### 3.3.1.1 Moving Pictures Expert Group (MPEG-2)

The Moving Pictures Expert Group is an ISO working group that is responsible for defining standards for the coded representation of digital audio and video. Since 1988, the group has established MPEG-1, MPEG-2, MPEG-4, MPEG-7 and MPEG-21. MPEG uses a lossy compression schema that sequentially stores changes from one picture and audio frame to the next. The most widely applied MPEG standard is MPEG-2. While MPEG-2 is based on MPEG-1 and is fully backward compatible, it produces much higher quality video and sound files. It has become the *De facto* standard for transmitting and storing digital video. The LAC recommends the use of MPEG-2 as the most appropriate format for the creation and preservation of digital video because of its status as an international standard, its market acceptance and penetration, and its apparent stability within the industry. During interchange, the MPEG-2 format must be MXF (Material eXchange Format) compliant.

For more information about the MPEG-2 standard, consult:
http://mpeg.telecomitalialab.com/standards/mpeg-2/mpeg-2.htm

For more information about MXF, consult:
http://www.broadcastpapers.com/sigdis/Snell&WilcoxMXF01.htm

### 3.3.2 Acceptable

#### 3.3.2.1 Audio Video Interleave (AVI)

Microsoft developed AVI for storing and playing audio and video data on a PC. The format is limited to a 320 x 240 video resolution and playback rate of 30 fps. AVI has become a *De facto* standard, but Microsoft has announced that it will soon drop support for the format. In the short-term, AVI files should be converted to a more stable format because its prospects for future support are not good. More information on AVI can be found at:
http://www.2dreamers.com/tutorials/John%20McGowan%27s%20AVI%20Overview.htm

#### 3.3.2.2 MPEG–4

MPEG–4 is built on the MPEG-1, MPEG-2 and Quicktime MOV (see below) standards. These files are designed for transmission over a narrow Internet bandwidth, making the file sizes smaller than other MPEG and Quicktime MOV file formats. MPEG-4 files can mix video with text, graphics, and 2-D and 3-D animation layers. The MPEG-4 standard has

yet to be adopted by many software developers and manufacturers. This is the reason that it is not a recommended LAC format. The MPEG-4 standard can be found at:
http://www.chiariglione.org/mpeg/standards/mpeg-4/mpeg-4.htm

#### 3.3.2.3 Quicktime (MOV)

The MOV file format was developed by Apple Computer to create, play and stream high-quality audio and video files on both Macintosh and Windows computers using the Quicktime software application. The format has been in use for over ten years and is fully backward compatible. The International Organization for Standardization chose the Quicktime format as the basis of the MPEG-4 standard. More information about Quicktime can be found at:
http://a992.g.akamai.net/7/992/51/c3264156652cee/www.apple.com/quicktime/products/qt/pdf/L29079A_QT6.3_DS.pdf

#### 3.3.2.4 Real Networks' RealVideo (RM)

RealVideo was the first streaming video format available on the World Wide Web. A RealVideo clip consists of two parts, a visual track that is encoded with RealVideo codecs (COmpression/DECompression) and an audio track encoded using RealAudio codecs. Both tracks are packaged in a RealVideo clip that uses the .RM file extension. RealVideo uses a lossy compression schema that reduces a video clip's size by lowering the frame rate or discards pixel data while recording the clip. More information on the RealVideo format can be found at:
http://www.realnetworks.com/resources/howto/audio_video/video.html

## 3.4 Documents - Textual

### 3.4.1 Recommended

#### 3.4.1.1 Extensible Markup Language (XML)

XML is a simple, flexible, and platform independent markup language derived from SGML (see below). It was designed to replace SGML because it is easier to understand and write code in XML while building applications for use on the World Wide Web. XML tags are fully extensible and user defined. They are used to describe the content of the text rather than its appearance. This allows for more efficient searching, but documentation of the tags is critical for one to be able to interpret a XML document. XML became a World Wide Web Consortium (W3C) recommendation in 1998 and it is now fully supported by all the leading software providers. A W3C 'recommendation' is a specification developed by a W3C working group and members of the consortium. A recommendation indicates the consortium members have reached a consensus that the specification is appropriate for widespread use. Since the use of XML is practiced at differing levels of technical maturity among GoC agencies and departments, the LAC is monitoring developments in the creation and use of domain specific XML schema definitions. The LAC will continue to monitor, evaluate and adopt specific XML formats as the schema definitions are developed, reviewed and approved by specific user communities. As these definitions are defined,

XML will become the LAC's preferred universal recommended standard for the interchange of digital information in the GoC.

*Versions:* 1.0 http://www.w3.org/TR/REC-xml
1.1 http://www.w3.org/TR/xml11/

#### 3.4.1.2 Extensible HyperText Markup Language (XHTML)

XHTML is a reformulation of HTML 4 (see below) as a XML application. XHTML 1.0 became a W3C recommendation in January 2000. XHTML 1.1 reformatted XHTML 1.0 into XHTML modules. This modularization provided the ability to extend and create subsets of XHTML, which made it easier to combine markup tags for vector graphics, multimedia, math, e-commerce and other applications. Version 1.1 became a W3C recommendation in May 2001. XHTML version 2.0 is currently being developed and will not be backwards compatible with previous versions. At the time of writing, version 2.0 cannot be considered stable. As a result, the LAC only recommends the use of XHTML versions 1.0 and 1.1. The LAC will continue to monitor the development of version 2.0.

*Versions:* 1.0 http://www.w3.org/TR/xhtml1/#xhtml
1.1 http://www.w3.org/TR/xhtml11
2.0 http://www.w3.org/TR/xhtml2

#### 3.4.1.3 HyperText Markup Language (HTML)

HTML is a simple markup system derived from SGML. It is used to create hypertext documents that are portable from one computer platform to another and it has become the standard format for producing documents for the World-Wide Web. Each HTML version contains a specific, non-extensible set of tags that are used to specify the appearance of the document being created. The LAC recommends that GoC departments and agencies produce HTML 4.01 documents rather than HTML 4.0 documents.

*Versions:* 2.0 http://www.w3.org/MarkUp/html-spec/html-spec_toc.html
3.0 http://www.w3.org/MarkUp/html3/CoverPage.html
3.2 http://www.w3.org/TR/REC-html32.html
4.0 http://www.w3.org/TR/html4
4.01 http://www.w3.org/TR/html401/

#### 3.4.1.4 Standard Generalized Markup Language (SGML) [ISO/IEC 8879:1986]

SGML is defined in international standard ISO 8879:1986. It is a markup language used for formally describing the structure and contents of documents. Tags in SGML are used to identify, name and describe relationships between data, so they can be managed and manipulated. SGML-based applications are platform independent and are used for a wide variety of functions. The SGML standard can be obtained from the ISO web site:
http://www.iso.org/iso/en/CatalogueDetailPage.CatalogueDetail?CSNUMBER=16387&ICS1=35&ICS2=240&ICS3=30

### 3.4.2 Acceptable

#### 3.4.2.1 Text Files (*.txt)

The LAC will accept plain text files that use the ISO/IEC 8859-1:1998 ASCII character set for encoding.

#### 3.4.2.2 Microsoft Word Document Format (.doc)

The .doc format is the native file format used to create documents in Microsoft Word. Microsoft Word is the most widely used word processing program throughout the world, thus the .doc format has become the *De facto* standard for the creation and distribution of textual documents.
*Versions:* 2.x, 4.0, 5.0, 5.1, 6.0/95, 97, 2000, and 2002

#### 3.4.2.3 Portable Document Format (PDF)

PDF is an open, *De facto* standard that was developed by Adobe for the electronic distribution of textually based documents in raster format. It is a widely used format that preserves all the fonts, formatting, graphics and colours contained in the original source document after its conversion to the PDF format. Although PDF is fully backwards compatible and platform independent, it is in the acceptable category because it is a proprietarily based solution. The Association for Suppliers of Printing, Publishing and Converting Technologies (NPES), and the Association for Information and Image Management International (AIIM International) are developing an international standard that defines the use of PDF for archiving and preserving documents. The format is known as PDF-Archive (PDF/A). The LAC is monitoring developments with respect to PDF/A becoming an ISO standard. Information about PDF/A can be found at:
http://www.aiim.org/documents/standards/SC2N226.pdf

*Versions:* 1.0, 1.1, 1.2, 1.3, and 1.4
http://partners.adobe.com/asn/developer/acrosdk/docs.html

#### 3.4.2.4 WordPerfect Document Format (.wpd)

The .wpd format is the native file format used to create textual documents in Corel WordPerfect. The WordPerfect software package is used extensively in GoC departments and in the private sector.
*Versions:* 1-5, 6.x, 7, 8, 9

## 3.5 Email

### 3.5.1 Recommended

#### 3.5.1.1 Multipurpose Internet Mail Extensions (MIME)

The Multipurpose Internet Mail Extensions format is an Internet standard that specifies how messages must be formatted so they can be exchanged between different email systems. MIME is very flexible and permits the inclusion of any type of file in an email

message. MIME messages can contain text, images, audio, video, or other application specific file types.
http://www.ietf.org/rfc/rfc2049.txt

## 3.6 Geospatial Data

### 3.6.1 Recommended

#### 3.6.1.1 Digital Line Graphs - Level 3 (DLG-3)

The DLG standard was originally developed by the U.S. Geological Survey (USGS) as a National Mapping Program (NMP) standard for the digital representation of many of the country's traditional 7.5-minute quadrangle cartographic paper maps. The format was created to define topological (i.e., spatial relationships between the data elements) vector-based line data such as roads, rivers and boundaries. Vector based data are constructed by using the point, line and polygon geometric primitive definitions. The DLG format is one of the more efficient, and widely recognized data formats used for the distribution of vector data. DLG-3 is gradually being replaced by the Spatial Data Transfer Standard (SDTS) interchange format (see below) in the United States Government. The DLG standards are available at:
http://rockyweb.cr.usgs.gov/nmpstds/dlgstds.html

#### 3.6.1.2 ESRI ARC/INFO Export (E00)

E00 is an interchange data format that was developed by Environmental Systems Research Institute (ESRI) to enable users to move data into and out of its geographic information system (GIS) software package known as ARC/INFO. A single E00 file describes a complete ARC/INFO coverage. An E00 file is actually an archive of smaller sub-files. There are two types of sub-files. Standard sub-files, which have fixed names and are comprised of a fixed data format that does not change from coverage to coverage. The second includes Info sub-files that contain user-defined attribute information. The E00 format is the LAC's preferred format for the transfer of digital maps in vector format.

#### 3.6.1.3 GeoTIFF

GeoTIFF files are TIFF images that have geographic coordinate data embedded as tags within the file. The geographic data are used to correctly position, orient and display the image in true geographic space. GeoTIFF makes use of a public tag structure that is platform independent. Most current GIS, CAD, image processing and desktop mapping applications can read GeoTIFF files that conform to the published specification. GeoTIFF files are the LAC's preferred format for the transfer of geographically referenced maps in raster format. The GeoTIFF specification can be found at:
http://www.remotesensing.org/geotiff/spec/geotiffhome.html

#### 3.6.1.4 Geography Markup Language (GML), Version 3

GML is an XML schema definition that is being developed by the Open GIS Consortium Inc. (OGC) for the transport and storage of geographic data. The format provides a

methodology for defining the geometry, topology, coordinate reference system, time and generalized attribute data that characterize the properties associated with geographic features. GML version 3 (GML3) conforms with the TC 211 ISO 191xx suite of standards for Geographic Information (see below). GML3 is also backward compatible with GML version 2.12. As GoC agencies and departments adopt application schemas using GML3, GML will become the LAC's preferred format for the interchange of geospatial data.

*Versions*: 1.0, 2.0, and 3.0
http://www.opengis.org/

#### 3.6.1.5 International Hydrographic Organization (IHO) S-57, Edition 3.1

The S-57: IHO Transfer Standard for Digital Hydrographic Data, Edition 3.1 was officially made available in November 2000. IHO S-57 is a standard that describes a data format for the transfer of digital hydrographic data. The standard is based on the ISO/IEC 8211:1994 specification for a data descriptive file for information exchange. The interchange standard is a media and content independent standard which allows users to name and describe data fields containing both character and binary data. Data structures in the S-57 format can be encoded in either binary or ASCII. The data structure is a tree with a finite number of levels: each file comprises records, each record fields, each field sub-fields.
*Versions:* 3.0, 3.1
http://www.iho.shom.fr/publicat/free/files/31Main.pdf

#### 3.6.1.6 TC 211 ISO 191xx Standards for Geographic Information

The Technical Committee 211 ISO 191xx suite of standards for digital geographic information is currently being defined. The standards will specify methods, tools and services for data management, processes for acquiring, processing, analyzing and presenting geographic information in electronic form and transferring data between different users and different systems. They will also provide a framework for the development of sector-specific applications using geographic data. Further information on the TC 211 ISO 191xx suite is available from the following web site:
http://www.isotc211.org

#### 3.6.1.7 Spatial Data Transfer Standard (SDTS)

SDTS is a United States Federal standard designed to support the transfer of different types of geographic and cartographic data. The standard defines the structure and content for spatial data to assist data transfer between different databases. SDTS is also known as the Federal Information Processing Standard (FIPS) 173.
http://mcmcweb.er.usgs.gov/sdts

### 3.6.2 Acceptable

#### 3.6.2.1 Canadian Council on Geomatics Interchange Format (CCOGIF)

This standard specifies the format for the exchange of digital spatial data among Canadian survey and mapping agencies. CCOGIF provides a national standard that preserves the accuracy and content of the exchanged information, and is machine and language independent.

*Versions:* 1.0, 1.1, 2.0, 2.1, 2.2, 2.3

http://www.cits.rncan.gc.ca/fich_ext/1/text/products/ntdb/ccogif.pdf

#### 3.6.2.2 CARIS ASCII

The CARIS software package is commonly used by international hydrographic agencies for the production of hydrographic charts. CARIS has a conversion utility that maps CARIS system files into an ASCII interchange format. The ASCII files can then be used for the transfer of data between different computer platforms that operate with incompatible character set representations. Although the LAC supports CARIS ASCII, it prefers that hydrographic data be transferred using the IHO S-57 interchange format. More information on CARIS ASCII can be found at:

www.caris.com

#### 3.6.2.3 CEOS Superstructure Format

The CEOS format consists of a generic component that defines the superstructure of the file coupled with a fixed record format that is adjusted for particular data types. The format only addresses the packaging scheme of the data and was designed to minimize the effort needed to read and write data from similar Earth observation sensors. No formal specification has been published for the CEOS format and because most agencies have developed their own software to create CEOS files, files created on one agency's software can often not be read by another agency. More information about CEOS is available from: http://wgiss.ceos.org/ceos.htm

#### 3.6.2.4 Digital Elevation Model (DEM)

A DEM data file consists of an array of terrain elevation samples for ground positions at regular intervals. It is used to create 3D graphics that display the slope, aspect and terrain profiles of a given area. The USGS DEM standard was recently altered to conform to the SDTS format and is available from:

http://rockyweb.cr.usgs.gov/nmpstds/demstds.html

#### 3.6.2.5 GeoVRML (Virtual Reality Modeling Language)

The GeoVRML file format is used to render geographic data using VRML, which is an ISO standard for representing 3D data over the Internet using a standard VRML97 browser. A geographic reference for the basic Cartesian coordinate system of VRML is implemented using the ISO standard Spatial Reference Model (SRM), which allows users to embed latitude/longitude or Universal Transverse Mercator (UTM) coordinates into VRML files.

GeoVRML is a "Recommended Practice" of the Web 3D consortium, but must be explored further before it becomes an LAC recommendation.

*Versions:* 1.0 http://www.geovrml.org/1.0/
1.1 http://www.geovrml.org/1.1/doc/

## 3.7 Structured Data – Databases and Spreadsheets

### 3.7.1 Recommended

#### 3.7.1.1 Flat File

All tabular data from legacy database and spreadsheet applications will be transferred to the LAC in an acceptable ASCII, EBCDIC or Unicode delimited flat file format. A flat file contains a sequentially arranged set of computer records that must be delimited with an end of record marker. Computer records are composed of a common logical grouping of data fields, which must contain an end of field delimiter for variable length records. Flat files are commonly used to transfer and import data files between users who do not use compatible software applications. For the future, the LAC will continue to monitor the use of XML schema definitions that are developed for the management of tabular data in database applications.

### 3.7.2 Acceptable

#### 3.7.2.1 dBase Format (DBF)

The dBase file format is widely used for the transfer of files between databases. The format was originally created for dBase database programs. The file header contains information about the record and is encoded in binary, while the record itself is encoded in ASCII. More information about dbf files can be found at:
http://www.e-bachmann.dk/computing/databases/xbase/dbf.html#DBF_STRUCT

## 3.8 Technical Drawings

### 3.8.1 Recommended

#### 3.8.1.1 Drawing Interchange File Format (DXF)

The DXF format is a tagged data representation of all the information contained in an AutoCAD® drawing file. DXF files enable the interchange of drawings between different CAD programs. The format is a tagged data representation of all the information contained in a drawing file. A number called a group code precedes each data element in the file. The group code's value indicates what type of data element follows. DXF files can be in either ASCII or binary formats. The LAC supports the ASCII format.

*Versions:* R2.05, R2.6, R9, R10, R11, R12, R13, R14, R2000, R2000i, R2002
http://usa.autodesk.com/adsk/servlet/item?siteID=123112&id=752569

## Bibliography

1. Adobe Developers Association. *TIFF Revision 6.0.* Mountain View, CA. (1992). http://partners.adobe.com/asn/developer/pdfs/tn/TIFF6.pdf

2. Bachmann, Erik. *Xbase Data File (*.dbf)* Denmark. (2003). http://www.e-bachmann.dk/computing/databases/xbase/dbf.html#DBF_STRUCT

3. Brooke, Simon. *XML Representation of Nautical Chart Data*. Scaffie Ltd. Auchencairn, Scotland. Retrieved June 003 from: http://www.weft.co.uk/library/xmlchart/documentation/overview-summary.html

4. Brooks, Alfred A. *Overview – ISO/IEC 8211:1994.* (1996). http://user.icx.net/~brooks/is8211.ovr

5. Brown, David et. al. *Management and Preservation of Geospatial Data.* Report written for the Ad-Hoc Committee on Archiving and Preserving Geospatial Data, GeoConnections, Policy Advisory Network Node, July. (2003).

6. California Digital Library. *Digital Image Format Standards.* (2001). http://www.cdlib.org/about/publications/CDLImageStd-2001.pdf

7. Canadian Council On Geomatics. *Standard File Exchange Format for Digital Spatial Data* – Version 2.3, (1994). http://www.cits.rncan.gc.ca/fich_ext/1/text/products/ntdb/ccogif.pdf

8. Cudlip, W. *Guidelines on Standard Formats and Data Description Languages Version 1.0.* Committee on Earth Observation Satellites. (1998). http://wgiss.ceos.org/documents/s&g/formguid.pdf

9. *F*ederal Ministry of the Interior. *SAGA: Standards and Architectures for eGovernment Applications*, KBSt Publication Series, Volume 56, February. Berlin, AG. (2003). http://www.kbst.bund.de/saga

10. *GIF Graphics Interchange Format.* CompuServe, Inc. Columbus, Oh. (1987). http://www.w3.org/Graphics/GIF/spec-gif87.txt

11. Interoperability Framework Coordination Group. *The HKSARG Interoperability Framework: Version 1.0. G*overnment of the Hong Kong Special Administrative Region Information Technology Services Department. November. (2002).

12. Hamilton, Eric. *JPEG File Interchange Format Version 1.02.* C-Cube Microsystems. Milpitas, Ca. (1992).
http://www.w3.org/Graphics/JPEG/jfif3.pdf

13. International Business Machines Corp., *IBM Character Data Representation Architecture, Reference and Registry, SC09-2190-00,* December. (1996).

*14.* International Organization for Standardization. *ISO/TC 211 Geographic Information/Geomatics Scope.* (2002).
http://www.isotc211.org/scope.htm#scope

15. ISO/TC171/SC2. *NWI Ballot for Document management – Long-term electronic preservation – Use of PDF (PDF/A).* International Organization for Standardization. Document N 226 E. April. (2003).
http://www.aiim.org/documents/standards/SC2N226.pdf

16. Lane, Tom. *JPEG Image Compression FAQ, part ½.* (1999)
http://www.faqs.org/faqs/jpeg-faq/part1/

17. Lim, Mark. *National Archives of Canada: Digital Media Formats Study.* 1514486 Ontario Inc. Contract No. 02011-2-0257. (2003).

18. McGowan, John F. *AVI Overview.* (1999)
http://www.2dreamers.com/tutorials/John%20McGowan%27s%20AVI%20Overview.htm

19. Moving Pictures Experts Group. *The MPEG Home Page.* Retrieved June 2003 from:
http://mpeg.telecomitalialab.com/

20. New Zealand E-government Unit. *New Zealand E-government Interoperability Framework (NZ e-GIF).* State Services Commission. Version 1.1. July. (2003).

21. Open GIS Consortium Inc. *OpenGIS Geography Markup Language (GML) Implementation Specification*. Document OGC 02-023r4, Version 3.0. editors, Simon Cox, et. al., January. (2003).

22. RealNetworks. *Video Production.* Retrieved June 2003 from:
www.realnetworks.com/resources/howto/audio_video/video.html

23. Quin, Liam. *XML Core Working Group Public Page* – Revision 1.24. World Wide Web Consortium. (2003).
http://www.w3.org/XML/Core/#Publications

24. Reddy, Martin and Iverson, Lee. *GeoVRML 1.1 Specification*. Web 3D Consortium. July. (2002).
http://www.geovrml.org/1.1/doc/

25. Ruth, Mike. *GeoTIFF FAQ Version 2.1.* (1999).
http://remotesensing.org/geotiff/faq.html

26. U.K. Office For Library and Information Networking (IKOLN). *NOF-digitise Technical Standards and Guidelines*. New Opportunities Fund, UKOLN, University of Bath in association with Resource: The Council for Museums, Archives & Libraries. Bath. Version Five: revised March. (2003).

27. U.S. General Service Administration. *Government Without Boundaries: A Management Approach to Intergovernmental Programs*. Office of Intergovernmental Solutions. May. (2002).

28. U.S. Geological Survey National Mapping Division. *Standards for Digital Line Graphs.* Department of the Interior. (1998).
http://rockyweb.cr.usgs.gov/nmpstds/acrodocs/dlg-3/1dlg0798.pdf

29. Usdin, B. Tommie et. al. *What is SGML?* Mulberry Technologies, Inc. Rockville, MD. (1997).
http://www.mulberrytech.com/papers/whatsgml.html