Library and Archives Canada

Guidelines for Computer File Types, Interchange Formats and Information Standards

Library and Archives Canada
Canadian Archives and Special Collections

Version 2.4.1
21 February 2008

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, Elizabeth Doyle and Ron Walker
Date	21 February 2008
Type	Text
Format	Microsoft Word 2003
Identifier	Version 2.4
Source	
Language	English
Relation	
Coverage	
Rights	Intellectual property rights – owned by Canada © Copyright – Her Majesty the Queen in Right of Canada - 2007

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 2.0	31 October 2007	Elizabeth Doyle	This version supersedes "Guidelines for Computer File Types, Interchange Formats and Information Standards", Version 1.0, 28 June 2004
Version 2.3	19 November 2007	Ron Walker	Revision to versions 2.0, 2.1 and 2.2 that were prepared by Ron Walker.
Version 2.4	10 December 2007	David Brown	Revisions to previous versions created by Elizabeth Doyle and Ron Walker.
Version 2.4.1	23 January 2008	David Brown	Draft release to Standards Working Group and the AV- Formats Working Group

Table of Contents

4 FILE TYPES AND INTERCHANGE FORMATS 19

1 Introduction

1.1 Purpose and Scope

This document identifies digital formats and standards which Library and Archives Canada (LAC) is recommending to facilitate the interoperability of digital information between LAC and its clients.

The file types and interchange formats cited in this document are intended to cover a number of data and digital information types; including audio, still images, moving images, textual documents, electronic publications, websites, email, geospatial data, databases, spreadsheets, and computer aided design (CAD) file formats that are used for the creation of architectural and technical drawings.

This document is not to be considered a complete list of available file formats; for a more complete registry please refer to the Library of Congress' Sustainability of Digital Formats website.

Although LAC has the technological capability to handle the entire set of file formats and standards identified in this document, those that appear are categorized into those that are "recommended" and those that are "acceptable" for use. Those identified as "recommended" are being promoted by LAC because they are best suited for the transfer and long-term preservation of digital information.

From an operational perspective, long-term access to digital information will be compromised unless policies, procedures and tools are created and implemented by creators to ensure its effective management and eventual preservation. Digital records are by their nature more fragile than paper records and permanent access to their content is more vulnerable to change or loss if they are not properly managed. 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.

1.2 Public and Private Records

Under the auspices of the Library and Archives of Canada Act, LAC is legislated to serve as the permanent repository for federal government and ministerial records that are of historical or archival value and to facilitate the management of information by federal government institutions. For public and private records, the term 'record' is defined in the

Act as "… any documentary material other than a publication, regardless of medium or form".[1]

The government of Canada is creating and storing terabytes of 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 Treasury Board of Canada Secretariat (TBS) develops information management (IM) policy and its implementation in the federal government is enhanced through guidance from LAC. To achieve the goal of preserving the records of interest to the nation and the government of Canada, LAC is promoting the formats and standards that are identified in this document for the exchange of digital information between federal government institutions and for the transfer of this information to LAC as historical or archival documentary heritage material.

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

Working in partnership with the library and archival communities, data producers need to standardize and adopt organizational policies and practices for their digital information holdings. The formats and standards identified here will serve to govern practices for the creation, use, retention, dissemination, preservation and disposition of digital information in federal government institutions; and will ensure its authenticity and integrity for as long as laws, regulations or government policies and directives require.

1.3 Publications

Also under the auspices of the Library and Archives of Canada Act, LAC is legislated to serve as the permanent repository for publications of the government of Canada. 'Publication' "… means any library matter that is made available in multiple copies or at multiple locations, whether without charge or otherwise, to the public generally or to qualifying members of the public by subscription or otherwise". [2]

Publications that are created by the federal government and private sector publishing houses may be made available through any medium (i.e., physical or virtual) and may be in any form, including analog or digital. Similar to the situation identified above for digital records, LAC acquires digital copies of published documentary heritage material

[1] Library and Archives of Canada Act
[2] Library and Archives of Canada Act

from the government of Canada and publishers for preservation purposes. To achieve this goal, LAC has begun the web harvesting of federal, provincial and territorial government websites, with the aim of ensuring that the sites that are collected and preserved as published documentary heritage collections, can be made available to Canadians for future consultation and research.

Preservation means ensuring that a publication survives long after copyright has expired. By acquiring an electronic publication from the originator as soon as it becomes published, LAC is assured of preserving the integrity of a publication as it was originally released.

1.4 The Digital Collection

The development of LAC's collection of digital materials is based on the broad collecting mandate established by the *Library and Archives of Canada Act* to acquire the published and unpublished documentary heritage of Canada. It is also governed by the more specific powers outlined in the legislation that relate to:

- The transfer of government and ministerial records of historical or archival value;
- The transfer of government records at risk;
- The powers that relate to the legal deposit of online publications;
- The representative sampling of the Internet; and,
- The provisions in the Legal Deposit regulations.

For further policy details and guidelines regarding the acquisition of digital materials please refer to LAC's Digital Collection Development Policy.

1.5 Concept

LAC has created this document to provide guidance on digital file types, interchange formats and information standards to public-sector departments and agencies in the Canadian federal government, as well as to private-sector publishers and donors. The adoption of these formats and standards will facilitate information exchange, provide a basis for the implementation of common 'information management' practices, and ensure the preservation of 'records and publications of value' for future generations of Canadians.

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 digital information for a period of five years before the information must be migrated to a new logical format. To preserve digital content and provide service to users and designated communities decades hence, custodians must be able to replicate the content on any type of media and have the ability to migrate and normalize the data in the face of changing technology standards. Migration and normalization procedures are costly to implement over time, 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 proposed file types, interchange formats and information standards that appear in this document, 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 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 to eliminate any potential reliance on the fate of any specific company recommendation. The formats appear in alphabetical order within the relevant areas.

The content for which a trusted digital repository takes long-term responsibility must not be protected by technical mechanisms such as encryption that may hinder the ability of custodians to preserve and provide access to the digital content. Nor are file formats that are inextricably bound to a particular physical carrier or devise considered to be suitable for long-term preservation. In addition, formats that have embedded capabilities that restrict the use of a file to protect intellectual property, that limit the use of a file for a period of time, that require users to enter a password to gain access, or formats that use other types of technical protection mechanisms that restrict user access and use, will not be accepted by LAC for preservation purposes. These factors apply to the way a file is used in a particular business context and this type of functionality must be neutralized prior to its transfer to the custody of LAC.

The embedding of meta-information into a file will not usually affect the use or the quality of the rendering of a file and it does not usually interfere with preservation functionality; for example, data that identifies rights-holders or the particular issuance of a work may appear in the header of a file. LAC encourages the use of meta-information because it useful for branding digital objects and indicates when a file was produced for a specific individual or entity. In addition, the information can be used to trace the movement of a file over time.

As indicated above, the physical medium upon which digital information is transferred and stored also plays a vital role in the preservation equation, but this issue will not be explicitly addressed in this document.

1.6 Updates

In order to maintain the currency of this document, the information presented herein will be reviewed and updated regularly. People are invited to comment on the contents; to direct comments please see the Enquires section.

1.7 Guidance

This policy should be read in conjunction with other relevant government of Canada legislation, policies and guidelines.

1.7.1 Federal Legislation

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

1.7.2 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
Directive on the Use of Official Languages in Electronic Communications
Privacy and Data Protection
Privacy Impact Assessment

1.7.3 Related Library and Archives of Canada Policies

Digital Collection Development Policy
Toward a Canadian Digital Information Strategy
Networked Electronic Publications Policy and Guidelines
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 Photographic Records in the Government of Canada
Federal Records Centres Users Guide
Understanding Electronic Records
Managing Internet and Intranet Information for Long-term Access and Accountability

1.8 Enquiries

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

Recordkeeping Liaison Centre
Library and Archives Canada
550 boulevard de la Cité
Gatineau, Quebec
K1A 0N4
Tel: 819-934-7519
Fax: 819-934-7534
Email: Centre.Liaison.Centre@lac-bac.gc.ca

1.9 Other Resources

Library of Congress
The Library of Congress has developed a website that serves as an inventory of information about current and emerging digital formats. It is being developed in association with the Global Digital Format Registry (GDFR) and the JSTOR/Harvard Object Validation Environment (JHOVE) projects.
http://www.digitalpreservation.gov/formats/index.shtml

Global Digital Format Registry (GDFR)
The purpose of the GDFR project is to provide a sustainable resource for managing format-critical representation information necessary for the long-term preservation of data.
http://hul.harvard.edu/formatregistry/

JSTOR/Harvard Object Validation Environment (JHOVE)
JSTOR, the 'Scholarly Journal Archive' and the Harvard University Library are collaborating on a project to develop an extensible framework for format identification and validation. Format identification is defined as: "… the process of determining the format to which a digital object conforms; in other words, it answers the question: 'I have a digital object; what format is it?'". Format validation is defined as: "… the process of

determining the level of compliance of a digital object to the specification for its purported format, e.g.: 'I have an object purportedly of format *F*; is it?'".
http://hul.harvard.edu/jhove/

Preservation and Long-term Access through Networked Services (Planets)
Planets is a four-year project co-funded by the European Union under the Sixth Framework Programme to address core digital preservation challenges. The project brings together European National Libraries and Archives, leading research institutions, and technology companies to address the challenge of preserving access to digital cultural and scientific knowledge.
http://www.planets-project.eu/

2 Presentation

2.1 Character Sets

Character sets refer to the system for encoding a sequence of characters in an eight bit byte, or octet. Traditionally, character sets and character encoding are considered to be synonymous terms.

2.1.1 Recommended

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

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 an 8-bit byte.

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]

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 LAC because it provides the default UCS encoding schema for HTML, SGML, XHTML and XML.

2.2 Markup Languages

ASCII, EBCDIC and Unicode define how each character in a specific computing environment will be created, but these character sets do not define how the characters and associated information will be displayed and organized in terms of their presentation. Markup languages enable one to dictate how information will appear by establishing structure to an object's layout information. Markup languages are used to set out the logical structure of a document through the use of standard syntax which define paragraphs or pages; for example, and which allows the structure of a document to be similarly rendered using multiple browsers and search engines, and specialized software.

The root code for all markup languages is SGML (Standard Generalized Markup Language) (ISO 8879:1985). It is a language that provides the specifications for an unlimited number of markup languages. All other markup languages, including HTML, XHTML, XML and Docbook, are derivatives of SGML.

2.2.1 Recommended

2.2.1.1 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.

An SGML document has three elements:

- The Declaration which describes the processing environment that is required;
- The Document Type Definition (DTD) which is a defined tag set that forms a template for describing the structure and content of a specific type of document; and,
- The Document stream itself.

Some relevant SGML DTDs are:

- **EAD** (Encoded Archival Description) which is a DTD for archival material;
- **US MARC DTD;**
- **HTML** (Hyper Text Markup Language);
- **XHTML** (eXtensible Hyper Text Markup Language);
- **XML** (eXtensible Markup Language).

2.2.1.2 Hyper Text Markup Language (HTML)

HTML is a simple markup language 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. LAC recommends that GoC departments and agencies produce HTML 4.01 documents rather than HTML 4.0 documents.

2.2.1.3 eXtensible Hyper Text Markup Language (XHTML)

XHTML is a reformulation of HTML 4 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, LAC only recommends the use of XHTML versions 1.0 and 1.1. LAC will continue to monitor the development of version 2.0.

2.2.1.4 eXtensible Markup Language (XML)

XML is a simple, flexible, and platform independent markup language that is considered to be a subset of SGML. It describes how you should format your tags, how you should document your definitions and how you will define your schema.

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.

Since the use of XML is practiced at differing levels of technical maturity among federal government agencies and departments, LAC is monitoring developments in the creation and use of domain specific XML schema definitions. 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 LAC's preferred universal recommended standard for the interchange of digital information in the federal government.

The Library of Congress has indicated that they would use XML as a preferred format if it was conformant to an appropriate standard or community agreed upon DTD or schema that can be used for technical validation.

3 Compression

In discussing file types, it is important to distinguish between file formats and compression algorithms. A codec (**co**mpressor/**dec**ompressor) is an algorithm that is installed on a PC and is capable of compressing or decompressing data in order for the file to be played by a particular software package. Codecs are usually applied to still imagery, moving imagery, or audio files, and can be further subdivided into lossy or lossless compression methods.

A ***lossless*** compression technique discards no information. It looks for more efficient ways to represent data, while making no compromises in accuracy. In contrast, ***lossy*** compression accepts some degradation in the data in order to achieve smaller file sizes. Because of this degradation in quality, lossy compression should be avoided for archival master images.

A codec is usually wrapped by a container, and most can be used to compress several different data formats. Conversely, most data formats can use different codecs. For example, an AVI file 'container' can use the DivX video codec, the MP3 audio codec or it could use the Indeo video codec or the PCM audio codec.

In some cases the codec and format are inherently linked, as is the case with proprietary software such as Real Player and Windows Media Player. In the case of MP3, AAC, or JPEG (and others) the codec can stand alone, or be wrapped in a different file format.

Most image compression techniques are independent of file formats, however some (JPEG, PNG) are inherently linked with a file format.

The most common compression algorithms for imagery consist of:

3.1.1 Run Length Encoding (RLE)

Run-length encoding is a simple lossless data compression method. It is well suited to standard palette-based images, but does not work well on continuous-tone images such as photographs.

RLE is used in fax machines, and is relatively efficient because most faxed documents are primarily composed of white space with occasional interruptions of black.

3.1.2 Lempel-Ziv (LZ)

The Lempel-Ziv (LZ) compression methods were developed by Abraham Lempel and Jacob Ziv in the late 1970's, and are among the most popular algorithms today for lossless storage.

3.1.3 Deflate

Deflate is a variation on LZ which is optimized for decompression speed and compression ratio, although compression can be slow. DEFLATE is used in PKZIP, gzip and PNG.

3.1.4 LZ-Reneau (LZR)

The LZR method forms the basis of the Zip compression scheme.

3.1.5 LZW Lempel-Ziv-Welch

LZW is a universal lossless data compression algorithm created by Abraham Lempel, Jacob Ziv, and Terry Welch in 1984, and was the first widely used universal data compression method. LZW is effective on 1-bit (monochrome) to 24-bit (True Colour) images.

Several file formats utilize LZW for compression, although it is most widely known as the compression algorithm for GIF files. More recently,an implementation of this algorithm is found within the popular Adobe Acrobat software to create PDF files.

3.1.6 CCIT Comité Consultatif International Téléphonique et Télégraphique

The CCITT, now known as ITU-T (ITU Telecommunication Standardization Sector) is the leading publisher of telecommunication technology, regulatory and standard information.

CCITT T.4 (commony referred to as Group 3 compression) is the universal protocol for sending fax documents across telephone lines. There are two levels of resolution:

203 by 98 and 203 by 196, and he compression protocol specifies a maximum transmission rate of 9 600 baud.

CCITT T.6 (commonly referred to as Group 4 compression) is a protocol for sending fax documents over ISDN networks. The Group 4 protocol supports images with a resolution of up to 400 dpi.

Both Group 3 and Group 4 are most commonly used by the TIFF file format. Several other ITU-T transmission standards exist; for more information go to: http://www.itu.int/home/index.html

3.1.7 Huffman Encoding

Huffman encoding is a lossless compression method developed in 1952 by David Huffman, and is one of the oldest and most established compression algorithms. Huffman encoding is designed to work best on images that have a lot of repetition, and is often used as a final compression stage in combination with more modern compression schemes such as JPEG, Deflate, and CCITT Group 3.

4 File Types and Interchange Formats

Generally speaking, file formats are specific patterns or structures which organize and define data. Some formats contain only one 'stream' of uncompressed data; others may contain codecs to compress the data, and others still may support several 'streams' of media.

The third type of format is called a 'container' or 'encapsulating' format. These formats can contain and support various types or layers of audio, video, still imagery, and their associated metadata. Each of these of these formats may be handled by different programs, processes, or hardware; but for the multimedia data stream to be interpreted properly, the information must be encapsulated together. The 'wrapper' refers to a particular way of storing and synchronizing the data content into a single file.

The most popular multimedia container formats are:

Proprietary:
AVI (standard Microsoft Windows container)
MOV (standard QuickTime container)
MP4 (standard container for the MPEG-4 multimedia portfolio)
ASF (standard container for Microsoft WMA and WMV)
RealMedia (standard container for RealVideo and RealAudio)

Open:
Ogg (standard container for Vorbis codecs)

Exclusive to Audio:
WAV (RIFF file format, widely used on Windows platform)
AIFF (IFF file format, widely used by the MacIntosh operating system).

Documents:
CGM (Computer Graphics Metafile)
PDF (Portable Document Format)

Please see specific sections in this document for information about these file types.

4.1 Digital Audio

Digital Audio is an audio signal encoded as bits of information. This information may be stored on any type of digital storage medium including a hard drive, USB flash drive, iPod, CD, DVD or tape.

4.1.1 Recommended

The recommended file formats are classified as being uncompressed file formats.

4.1.1.1 Audio Characteristics

Accurate digital sound quality (high fidelity) depends on the range of sound which is sampled, the rate at which it is sampled, and the various conversions that occur in any sound reproduction system.

Fidelity is a factor associated with sound waveforms, and refers to the degree of accuracy in which digital content may be reproduced. In general, uncompressed or lossless compressed data offers the highest fidelity; however, lossy compression based on human perception of sound provides a high level of fidelity in normal playback conditions.

Fidelity characteristics should be used as the primary consideration for audio quality. The choice of file format is considered to be a secondary factor.

Preferred fidelity: **24 bit at 48 kHz or higher**.
Minimum fidelity: **16 bit at 44.1 kHz**.

When determining the fidelity characteristics (bit rate, sample rate) of a file within Windows Explorer, right-click the audio file name, and then select Properties>Summary>Advanced.

4.1.1.2 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. The format was developed by Apple Computer in 1988 and is the standard audio format for Macintosh computers. In addition to Macintosh computers, the AIFF format is platform independent and is supported by all audio editing and player software.

AIFF 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.

4.1.1.3 Broadcast Wave Format (BWF)

The European Broadcast Union (EBU) introduced BWF in 1996 to allow files to be exchanged between digital audio workstations during radio and television productions. It is now used in every aspect of professional audio. Based on Microsoft's and IBM's WAV format, BWF can carry PCM (Pulse Code Modulation) or MPEG encoded audio which can be enhanced with metadata that describe information about the originator, date and coding history. A BWF file has WAV as its file name extension and is read by software that supports WAV files. The International Association of Sound and Audiovisual Archives (IASA) recommend its use as an archival audio file format, and LAC has recently switched to this from WAV.

4.1.1.4 Waveform Audio Format (WAV)

Microsoft and IBM developed the WAV format jointly. Support for WAV files was built into the Windows operating system, making it the standard for sound on PCs. The format supports many bit resolutions, sample rates, audio channels and a number of lossless compression methods; however WAV files are very rarely compressed.

The most common WAV format contains uncompressed audio in the Pulse Code Modulation (PCM) format. PCM audio is the standard audio file format for recording CDs at 44,100 samples per second, 16 bits per sample. Since PCM uses an uncompressed, lossless storage method, the WAV format maintains maximum audio quality.

As a long-standing digital audio format, WAV remains the *De facto* standard for lossless PC audio files in use today.

4.1.2 Acceptable

4.1.2.1 MPEG-1 layer-3, MPEG-2 layer3 (MP3)

Preferred fidelity: **192 Kilo bits per second (Kbps) @ 44.1 Kilo Hertz (KHz)**
Minimum fidelity: ?

MP3 is a lossy codec, and is the most widely-used digital audio format, today. It works by eliminating inaudible frequencies in combination with a high compression rate (rates of 10:1 are possible).

The primary advantage of MP3 is its universality; unlike most other file formats, just about every digital music player and player program can handle the MP3 format for music.

4.1.2.2 RealAudio v.10 (RA, RAM)

RealAudio is a proprietary lossless container file format for audio that was developed by RealNetworks. RealAudio was developed as a streaming media product, meaning that it can be played while it is being downloaded, and it has become a *De facto*

standard for network audio. RealAudio 10 incorporates the MPEG-4 Advanced Audio Coding (AAC) codec at bit rates of 128 Kbps.

4.1.2.3 MPEG-4 AAC - Advanced Audio Coding (AAC)

AAC is a lossy codec that was created by Fraunhofer-Gesellschaft as the next generation MP3 codec. Due to advances in the technology, AAC files encoded at 96 Kbps sound slightly better than a MP3 file that is encoded at 128 kbps.

The format comes in 2 varieties: MPEG-2 AAC and MPEG-4 AAC and both are referred to as AAC. MPEG-2 AAC is used for several purposes. It is part of the specification for the DVD-Audio Recordable (DVD-AR) format and it can also be used for streaming and downloading online audio.

MPEG-4 AAC is a newer specification that includes more capabilities. It serves as a multimedia container format and delivers higher-quality sound than MPEG-2. Its popularity is currently maintained by it being the default codec for Apple's iTunes; the media player which powers iPod which is currently the most popular digital audio player on the market. This format is used by Apple because it supports digital rights management (DRM).

NOTE: The m4p extension indicates that a file is protected, whereas the .m4a extension indicates that a file is unprotected.

4.1.2.4 Windows Media Audio (WMA)

This is a proprietary compressed audio file format developed by Microsoft and is similar to the popular MP3 format. With the introduction of WMA Pro and Apple's iTunes Music Store, WMA has positioned itself as a competitor to the AAC format used by Apple and is part of Microsoft's Windows Media framework.

4.1.2.5 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.

4.1.3 Formats Under Investigation

4.1.3.1 AUdio file (AU)

AU is the standard audio file format used by Sun, Unix and Java. The audio in AU files can be PCM or be compressed with the ulaw, alaw or G729 codec algorithms.

4.1.3.2 Vorbis (Ogg)

Ogg is the name of Xiph.org's container format for audio, video, and metadata. This format supports a variety of codecs, but it is most commonly associated with Vorbis. Ogg Vorbis files are often compared to MP3 files in terms of quality.

The Ogg Vorbis format has proven popular among supporters of open source code. They argue that its higher fidelity and completely free distribution make it an excellent replacement for patented and restricted formats such as MP3.

Many mainstream video game titles store game audio as Vorbis files. Popular software players support Ogg Vorbis playback either natively or through an external plug-in. It is also suitable for online streaming, and is currently being used by Radio France, Virgin Radio, and it is currently being tested by CBC radio.

4.2 Digital Video

Generally, digital video consists of bits of data which produce and display pixels that make up the frames of a video sequence. The higher the frame rate the better the motion, and the higher the bits per pixel, the better the colour quality.

This information may be stored on any digital storage device including a hard drive, USB flash drive, CD, DVD or any type of tape media. The data may also be compressed via a video codec. While audio codecs are either lossy or lossless, video codecs are normally lossy. Since there is such a large amount of data held in a high-resolution video file, it is impractical to use lossless compression.

Unlike analog video which is typically stored on tape, subject to wear and where each copy is slightly inferior to the original, it is possible to copy multiple generations of digital video files with no degradation in quality.

4.2.1 Recommended

4.2.1.1 Digital Video Characteristics

Preferred fidelity: **192 Kbps @ 44.1 KHz**
Minimum fidelity: ?

4.2.1.2 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. MPEG uses a lossy compression schema that sequentially stores changes from one picture and audio frame to the next. Currently, there are three major MPEG video standards, MPEG-1, MPEG-2 and MPEG-4.

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 offers resolutions of 720 x 480 and 1280 x 720 at 60 frames per second. It has become the *De facto* standard for transmitting and storing digital video, and is the standard DVD file format. MPEG-2 is formally known as ISO/IEC 13818-2 and as International Telecommunication Union (ITU) Telecommunication (ITU-T) Recommendation H.262.

LAC prefers the use of MPEG-2 as the 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 and preferably based on the 4:2:2 chroma profile.

4.2.2 Acceptable

4.2.2.1 Audio Video Interleave (AVI)

Microsoft developed AVI as a multimedia container format that conforms to RIFF (Resource Interchange File Format) specifications. An AVI file may carry audio or visual data in almost any compression scheme, including: Full Frames (Uncompressed), Intel Real Time Video, Indeo, Cinepak, Motion JPEG, Editable MPEG, VDOWave, ClearVideo / RealVideo, QPEG, or MPEG-4.

AVI is considered by most to be an outdated container format, yet despite its limitations it remains popular among file-sharing communities.

4.2.2.2 Moving Pictures Expert Group (MPEG–4)

MPEG-4 was introduced in 1998 and is built on the MPEG-1, MPEG-2 and Quicktime MOV (see below) standards that are based upon ISO/IEC 14496. Similar to MPEG-2, the uses for the MPEG-4 standard are web streaming, and audio and video file distribution. MPEG–4 files are designed for transmission over a narrow Internet bandwidth because the file sizes are smaller than the other MPEG formats. MPEG-4 is still a developing standard and is divided into a number of parts. Companies promoting MPEG-4 compatibility do not always clearly state which "part" of the standard they are supporting; as a result, the format is not a recommended LAC format.

4.2.2.3 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.

4.2.2.4 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 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.

4.2.2.5 Windows Media Video (WMV)

WMV is short for Windows Media Video, a generic name for video codecs developed by Microsoft, as non-standard versions of MPEG-4. As of March 2006, WMV 9 was officially declared an independent Society of Motion Picture and Television Engineers (SMPTE) standard, and it is now considered to be a unique independent codec.

The WMV 9 codec is similar to ASF (Advanced Systems Format), and is often packed into ASF containers, with the resulting file named .wmv or .asf. WMV compressed files can also be put in AVI containers.

The WMV container format usually combines the WMV 9 video codec with the Windows Media Audio stream. The advantage of this format it that it allows the compression of large video files while retaining considerably high quality.

4.2.3 Formats Under Investigation

4.2.3.1 Theora (Ogg)

Theora is an open video codec developed by the Xiph.org Foundation as part of their Ogg project. This project aims to integrate On2's VP3 video codec, the Ogg Vorbis audio codec and the Ogg multimedia container format into a multimedia solution that can compete with MPEG-4. Theora is derived directly from On2's VP3 codec; currently the two are nearly identical, varying only in framing headers, but Theora will diverge and improve from the main VP3 development lineage as time progresses.

4.2.3.2 Motion JPEG 2000 (MJ2, MJP2)

Motion JPEG 2000 is a file format for the storage of multiple JPEG images that appear to be in motion. Support for associated audio is also included. The format is defined as Part 3 of the JPEG 2000 standard.

MJ2 does not involve inter-frame coding: each frame is coded independently using JPEG 2000. Expected applications include:

- storing video clips taken using digital still cameras;
- high-quality frame-based video recording and editing;
- digital cinema; and,

- medical and satellite imagery.

MJ2, originally defined in ISO/IEC 15444-3:2002 as a standalone document has now been expressed by ISO/IEC 15444-3:2002/Amd 2:2003 in terms of the ISO Base format, ISO/IEC 15444-12.

According to The Society for Imaging Science and Technology, Motion JPEG 2000 has the potential to be a format for long-term video preservation, and the U.S. National Library of Medicine is promoting MJ2 as their format for long-term video preservation.

4.3 Digital Still Imagery

Digital image files are made up of either raster (pixel) or vector (geometric) data, and a variety of file formats exist for either type of image. This information may be stored on any type of digital storage device including a hard drive, USB flash drive, CD, DVD or tape media.

4.3.1 Recommended

4.3.1.1 Imagery Characteristics

We need to define this……

4.3.1.2 Joint Photographic Experts Group (JPEG)

JPEG refers to both a compression and a file format. The JPEG compression is a standardized lossy image compression designed for compressing full-colour and grayscale images. The International Organization for Standardizaton (ISO) standardized the JPEG compression format in 1990.

It uses a 24-bit colour depth, and the compression 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. This algorithm can be used to compress data within several different file formats, including JFIF, TIFF, or PDF to name a few.

The JPEG file fomat is **JFIF** (**JPEG Image File Format**), but uses the JPG extension. When most people refer to JPEG, JFIF is the file format to which they are referring. JFIF is fully compliant with the JPEG standard.

4.3.1.3 Joint Photographic Experts Group JPEG 2000 (JP2)

JPEG 2000, Part 1 (the core system, .jp2), is an open, published international ISO standard (ISO 15444). It offers both lossless and lossy compression and provides superior image quality with more efficient compression than JPEG.

JPEG 2000, Part 6 (for mixed raster content, .jpm), is also a published international ISO standard (ISO 15444/6). It is aimed at compressing scanned colour documents containing both bi-tonal elements as well as images.

Compared to JPEG (which is limited strictly to the RGB colour space), JPEG 2000 compression supports grayscale, RGB and CMYK colour models in up to16-bit colour; in addition to alpha and spot color channels. It can also contain XML compliant metadata.

It should be noted that Microsoft is promoting it's own file format which is reported to be comparable to JPEG 2000; HD Photo (.hdp, .wdp). HD Photo is the new name for Windows Media Photo, and is the native image file format in the Windows Vista operating system.

4.3.1.4 Portable Network Graphics (PNG)

PNG is an open lossless extensible file format that was designed to provide a patent-free, high quality replacement for the GIF and TIFF file formats. It supports the capability of storing up to 16-bits (gray-scale) or 48-bits (truecolour) per pixel, and up to 16-bits of alpha data. It also handles the progressive display of image data and the storage of gamma, transparency and metadata.

4.3.1.5 Tagged Image File Format (TIFF)

TIFF is a bitmapped image format developed by Aldus (now part of Adobe) in 1986. TIFF files may or may not be compressed, and are extensible and portable. They do not favour a particular computer operating system, compiler or processor.

Baseline TIFF images support monochrome, gray-scale, palette (i.e., indexed), and RGB (i.e., true colour) colour spaces. A common extension of TIFF also allows for CMYK images.

TIFF Uncompressed baseline* RGB TIFF v.6.0 is LAC's preferred standard for describing and storing raster image data from scanners, faxes and digital photography applications, however because this format can come in a range of types, it is important to specifically note that our preference is for **v.6.0 uncompressed baseline RGB TIFF** files.

* for baseline specs see http://www.awaresystems.be/imaging/tiff/tifftags/baseline.html

The TIFF format uses a 32-bit colour depth, and is therefore limited to a maximum file size of 4 gigabytes.

4.3.2 Acceptable

4.3.2.1 Digital Imaging and Communications in Medicine (DICOM v.3.0)

DICOM is a universal image format standard used by virtually all modern ultrasound devices, X-ray photography systems and computer tomographs (CT scans) for storage and transmission of medical images.

DICOM standards are designed to achieve compatibility and improve workflow efficiency between imaging and other information systems in healthcare environments, and are maintained by the DICOM Standards Committee.

The DICOM file format is supported by special software shipped together with corresponding medical equipment, however the DICOM standard is also supported by most devices which allow data exchange regardless of equipment or examination type.

4.3.2.2 Encapsulated PostScript (EPS)

An EPS file is a Post Script language program describing the appearance of a single page, and may contain any combination of text, graphics and images.

LAC receives EPS files from Canada Post, as it is their most common file format used in relation to stamp design.

4.3.2.3 Graphics Interchange Format (GIF)

GIF is primarily an exchange and storage format, although it is supported by many applications. 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. GIF images are lossless and compressed using the LZW scheme.

GIF files are limited to 256-bit images, so the format is more suited to monochrome logos and graphics than for colour photographs.

4.3.3 Formats Under Investigation

4.3.3.1 Basic Image Interchange Format (BIIF)

This interchange format is based on the National Imagery Transmission Format Standard (NITFS) developed by the US Department of Defence adopted by NATO. BIIF is the basis for a new standards activity within ISO/IEC JTC1/SC24 (Computer Graphics and Image Processing) to add a new Part 5 to the International Standard for Image Processing and Interchange (IPI) (ISO 12087) that is based on BIIF.

4.3.3.2 Enhanced Compressed Wavelet (ECW)

ECW is a wavelet based image compression format that was designed to handle lossy compression and local or internet access to very large size RGB or greyscale images at 95% or so compression rates (20:1).

4.3.3.3 PDF For Long-Term Preservation (PDF/A)

PDF/A is a subset of PDF (Portable Document Format) for the long-term archiving of electronic documents. There are two levels of compliance to the PDF/A specification, Part 1 and Part 2.

Part 1 (PDF/A-1) is based on the PDF Reference Version 1.4 from Adobe Systems Inc. (as implemented in Adobe Acrobat 5), and is a certified ISO standard: ISO 19005-1: Document Management - Electronic document file format for long term preservation - Part 1: Use of PDF 1.4.

Part 2 (PDF/A-2) will be a new version of the PDF/A standard, and will identify a profile that will ensure that documents are self-contained, and can be reproduced in the future. The format will be based on the PDF Reference Version 1.6, or newer. According to the ISO Working Group, the release of this new version is not expected before 2008.

In December 2006, Berlin-based software manufacturer Callas released an Adobe Acrobat plug-in which allows users to check existing PDF files for compliance with the Part 1 ISO standard and also provides the capability for users to convert them to PDF/A files.

4.4 Text

There are two general types of text; plain and formatted. Plain text files contain encoded ASCII or Unicode data that has no formatting or layout code to influence the presentation of the data. These files can be opened and read by any software package that is capable of interpreting and displaying the numeric and alpha-numeric data values that are contained in a file.

Formatted text files such as rich text (rtf) contain encoded ASCII data and format definitions that display the information in a defined pattern. Formatted text files are primarily created using word processing software such as WordPerfect and Microsoft Word. Markup languages such as HTML and XML are also considered to be formatted text.

Generally speaking, textual file formats can be subdivided into three categories: Word Processing, Structural Markup, and Page Layout.

The current de-facto formatted document file format is Microsoft Word, which uses the .doc extension. However, the OpenDocument format (ODF) is slowly gaining recognition, and may emerge as the preferred file format for the creation of office documents, in the future (see Formats Under Investigation for further details regarding the ODF format).

Library and Archives Canada does not recommend the acquisition of proprietary binary formats that are created by various word processing software applications such as Microsoft Word and WordPerfect, or desktop-publishing applications such as Quark. However, the institution recognizes that this may be impossible to apply in all situations. Furthermore, the Library of Congress recommends that text documents that are created by the most popular word processing applications be converted to the PDF format (preferably PDF/A), or to a non-proprietary format such as OpenOffice, which is XML-based.

4.4.1 Recommended

4.4.1.1 HyperText Markup Language (HTML)

See Markup Languages for details on HTML.

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

See Markup Languages for details on SGML.

4.4.1.3 Extensible Markup Language (XML)

See Markup Languages for details on XML.

4.4.1.4 Extensible HyperText Markup Language (XHTML)

See Markup Languages for details on XHTML.

4.4.1.5 Multipurpose Internet Mail Extensions (MIME)

The MIME format is an Internet standard that specifies how messages must be formatted so they can be exchanged between different email systems. MIME allows email messages to contain:

- Multiple objects in a single message.
- Text having unlimited line length or overall length.
- Character sets other than ASCII, allowing non-English language messages.
- Multi-font messages.
- Binary or application specific files.
- Images, Audio, Video and multi-media messages.

4.4.1.6 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. PDF is fully backwards compatible and platform independent.

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). LAC is monitoring developments with respect to PDF/A becoming an ISO standard.

4.4.1.7 Plain Text (TXT)

LAC will accept plain text files that use the ISO/IEC 8859-1:1998 ASCII character set for encoding. Plain text files use the extension .txt and contain ASCII-encoded text with no formatting or layout information. They can be opened and read by any program that reads text.

4.4.1.8 Rich Text Format (RTF)

The RTF Specification provides a format for text and graphics interchange that can be used with different output devices, operating environments, and operating systems. RTF uses the ANSI, PC-8, Macintosh, or IBM PC character set to control the representation and formatting of a document, both on the screen and in print. With the RTF Specification, documents created under different operating systems and with different software applications can be transferred between those operating systems and applications. RTF files created in Word 6.0 (and later) for the Macintosh and Power Macintosh have a file type of "RTF."

4.4.2 Acceptable

4.4.2.1 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.

4.4.2.2 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.

4.4.3 Formats Under Investigation:

4.4.3.1 Open Document Format (ODF)

The OASIS (Organization for the Advancement of Structured Information Standards) OpenDocument Format (ODF) is an XML-based file format for office documents, including, textual documents, spreadsheets, drawings and presentations.

ODF is based on ISO standard ISO/IEC 26300:2006, which is considered to be an open format. Any software maker can use the associated specification and develop software applications that are based on the standard. The ODF standard has been chosen by seven nations (e.g., Brazil, France, Germany, Belgium, Croatia, Norway and Demark) for the creation and management of their respective government business documents.

ODF offers alternative products to proprietary software solutions that have been developed by various software manufacturers – see below.

Product	**Proprietary software**	**OpenDocument equivalent**
Word processor	MS Word	Writer
Spreadsheet	MS Excel	Calc
Presentation	MS PowerPoint	Impress
Database	MS Access	Base
Vector Graphics Editor	CorelDraw	Draw

Microsoft has developed an equivalent format to ODF, called Open XML (OOXML), which is not totally compliant with the ISO/IEC 26300:2006 standard. Microsoft will be using their version of the ODF format for the creation of office documents that use the Office 2007 suite of tools in the Windows Vista operating system environment.

4.5 Structured Data (Databases and Spreadsheets)

Structured data refers to data that resides in fixed fields within a record or file. (e.g. spreadsheets and relational databases.

4.5.1 Recommended

4.5.1.1 Flat File

All tabular data from legacy database and spreadsheet applications will be transferred to 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, LAC will continue to monitor the use of XML schema definitions that are developed for the management of tabular data in database applications.

4.5.2 Acceptable

4.5.2.1 DBF dBase Format

The dBase file format is widely used for the transfer of files between databases. The format was originally created for dBase database applications. The file header contains information about the record and is encoded in binary, while the record itself is encoded in ASCII.

4.6 Web Formats

4.6.1 Recommended

4.6.1.1 Internet Archive Format (ARC)

This format was created by the Internet Archive as a method for combining multiple digital resources into a single archival file. It is used to store 'web crawls' as sequences of content blocks harvested from the World Wide Web.

4.6.1.2 Web ARChive Format (WARC)

WARC is the next generation of the ARC format. WARC generalizes the older format to better support the harvesting, access, and exchange needs of archiving organizations.

Besides the primary content currently recorded, the revision accommodates related secondary content, such as assigned metadata, abbreviated duplicate detection events, and later-date transformations.

4.7 Geospatial

Geospatial data consists of spatially defined geographic data that are analyzed through the use of Geographic Information System (GIS) software, image processing systems, or similar types of modeling software and technology. Geographic and spatial data comes in a variety of formats, some of which can be used in GIS software

applications. The most popular formats that are used in the Canadian federal government appear below.

4.7.1 Recommended

4.7.1.1 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.

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.

4.7.1.2 Environmental Systems Research Institute (ESRI) Export Format - (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. 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.

4.7.1.3 Environmental Systems Research Institute (ESRI) Shape File Format (SHP)

ESRI introduced the Shapefile to provide GIS users with a simple and effective means to disseminate geospatial information, as an alternative to the E00 export file format. As a result, the Shapefile is becoming the leading de facto standard for geospatial data exchange and desktop GIS applications. The openly published Shapefile format is based upon a nonproprietary geospatial data structure.

4.7.1.4 GeoTIFF

GeoTIFF files are TIFF images that have geographic coordinate data embedded as The GeoTIFF specification defines a set of TIFF tags which describe cartographic information associated with TIFF imagery including projections, coordinate systems, ellipsoids, datums, and anything else required to establish the spatial reference of an image.

GeoTIFF files make use of a public tag structure that is platform independent, and is fully compliant with TIFF 6.0. Image content includes satellite imaging systems, aerial

photography, scanned maps, digital elevation models, or the results of geographic analyses.

GeoTIFF files are LAC's preferred format for the transfer of geographically referenced maps in raster format.

4.7.1.5 Geography Markup Language (GML)

GML is an XML Document Type Definition (DTD) that has been developed by the Open Geospatial Consortium (OGC) for the transport and storage of geographic data. The format provides a methodology for defining the geometry, topology, coordinates reference system, time and generalized attribute data that characterize the properties associated with geographic features.

The OGC is an international standards organization whose members maintain the *Geography Markup Language* standard. The OGC coordinates with the ISO TC 211 standards organization to maintain consistency between OGC and ISO standards work. GML has now been adopted as an ISO standard and conforms to the TC 211 ISO 19100 suite of standards for Geographic Information (see below). As federal government departments and agencies adopt GML, GML will become LAC's preferred format for the interchange of geospatial data.

4.7.1.6 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.

4.7.1.7 TC 211 ISO 19100 Suite of Standards for Geographic Information

ISO Technical Committee 211 is developing the 19100 suite of standards for geographic information and geomatics. The standard specifies the methods, tools and services for data management; processes for acquiring, processing, analyzing and presenting geographic information in digital form; and, transferring data between different users and systems.

4.7.1.8 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.

4.7.1.9 Spatial Archive Interchange Format (SAIF)

The Spatial Archive and Interchange Format (SAIF) features a powerful object-oriented data model described in an easy-to-use data definition language called Class Syntax Notation (CSN). SAIF is the standard archive and interchange format for geographic data in the province of British Columbia. SAIF was developed to address both data in-terchange and data archival issues.1As a result, SAIF is an excellent format for storing geographic data in a vendor-neutral manner. The Feature Manipulation Engine (FME) enables data stored in SAIF to be easily translated to any of the popular vendor for-mats.

4.7.2 Acceptable

4.7.2.1 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.

4.7.2.2 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.

4.7.2.3 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.

4.7.2.4 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.

4.7.2.5 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.

4.7.3 Formats Under Investigation:

4.7.3.1 Binary Universal Format Representation (BUFR)

BUFR is a binary code desiged to represent any meteorological data. The final file (this data set) is a monolithic BUFR file known as PREPBUFR. All data types used by the model are present, including upper air, marine surface, aircraft reports (ACARS, AIRCAR), land surface,etc. Major parameters include location of data and particulars about the instruments; wind and turbulence information; temperature information including dry bulb, wet bulb, dew point, soil, etc.hygrographic and hydrological information including precipitation, snow and snow depth, river stage, relative humidity, and others; radiation and radiance; ozone and air mass; synoptic features; present weather;oceanographic data including wave and swell and others; dispersal and transport; and radiological elements.

4.7.3.2 Network Common Data Form (NetCDF)

NetCDF is a machine-independent, self-describing, binary data format standard for exchanging scientific data. The project homepage is hosted by the Unidata program at the University Corporation for Atmospheric Research (UCAR). They are also the chief source of netCDF software, standards development, updates etc. The format is an open standard.

The project is actively supported, currently at version 3, and with plans underway for a version 4. It is planned to merge the project in version 4 with the HDF5 data file format, giving an alternative netCDF interface to HDF5 files.

The format was originally based on the conceptual model of the NASA CDF but has since diverged and is not compatible with it.

4.8 Technical Drawings

4.8.1 Recommended

4.8.1.1 Drawing Interchange File Format/Data eXchange 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 software applications. DXF was originally released in 1982 as part of AutoCAD 1.0, and the specification was intended to provide programmers the data model for the AutoCad native file format known as DWG. DXF files can be in either ASCII or binary formats. LAC supports the ASCII format.

4.8.2 Acceptable

4.8.2.1 Computer Graphics Metafile (CGM)

Although CGM is not widely supported and has been supplanted by other formats, it is still prevalent in engineering, aviation, and other technical applications. CGM is a file format for 2D vector graphics, raster graphics, and text, and is defined by ISO/IEC 8632.

5 Definitions

5.1 Raster Graphics (Bitmap Images)

A raster image is comprised of bits of information representing uniquely valued pixels in the form of a grid. Image resolution is measured by pixels per inch (PPI); however the printing abbreviation DPI (dots per inch) is also commonly used to describe image resolution. All digital photographs, regardless of file type, are raster images.

150 PPI

50 PPI

The more pixels there are in relation to the area, the higher the resolution. The higher the resolution, the sharper the image is and the larger the file. The picture to the left has a ppi of 150; the picture below has a resolution of 50 ppi.

Digital image resolution is greatly misunderstood. Digital images themselves have no size other than the number of pixels they contain. The image only has real dimensions (inches or cm) when it is in an analogue form before digitization, or after it has been printed.

There are two basic measures for digital imagery characteristics:
Spatial resolution – capturing detail (PPI); and,
Tonal resolution – colour, bit depth and dynamic range.

Generally, the higher the PPI and the larger the bit depth, the more accurate the image will be to it's original colour. Black and white images are not characterized by colour resolution. They are comprised of brightness values that represent 256 different shades of gray.

Colour Depth
The number of colours available in a digital image is determined by the number of bits assigned to each pixel. The more bits per pixel, the more colours can be displayed.

Colour Depth	Number of Colours Visible
1 bit (monochrome)	2
4 bit	16
8 bit (indexed colour)	256
24 bit (true colour)	16,777,216

Common "colour resolutions" are 1 bit per pixel, for solid black-and-white nonrealistic images; 8 bits per pixel for grayscale images, nonrealistic colour images, and coarse realistic images; and 24 bits per pixel, for "photographic quality" realistic images. 48 bits per pixel is in increasing use for ultrahigh quality images.

Grayscale images have a maximum colour depth of 8 bits. This is because when defining shades of gray in terms of RGB, each of the 3 red, green and blue components must be equal (i.e. R=192 G=192 B=192, or R=128 G=128 B=128). Since these three components must be equal, there are only 256 possible combinations, which equals 8 bits of colour.

Indexed colour images are limited to a maximum of 256 colours (8-bit), which can be any 256 colours from the set of 16.7 million 24 bit colours. Each image file contains it's own palette which provides a reference index number used by the computer to identify each colour.

GIF, TIFF and PNG files use indexed colour, but only GIF requires indexed colours. TIFF and PNG can be stored as indexed colour or true colour.

Images are known as 'true colour' when each pixel is defined in terms of its actual RGB or CMYK values. In the RGB model, each pixel in a a true colour image has 256 possible values for each of it's red, green or blue components, or in the CMYK model, it's cyan, magenta, yellow and black components.

Because there are 256 possible values for each RGB or CMYK component, an RGB true colour image would have a 24-bit colour depth and a CMYK true colour image would have a 32-bit colour depth.

5.2 Vector Graphics

Page Description Languages (PDLs) describe the contents of a printed page (layout, font, graphics) to a display device or printer. The de facto standard PDL is PostScript(PS), a programming language primarily developed to work with text, but is

capable of containing raster and vector images. Recently however, newer forms of PostScript like EPS (Encapsulated PostScript) and PDF (Portable Document Format) are replacing PostScript because of their ability to exchange graphics, and according to TASI are "blurring the distinction between PDLs and metafiles".[3]

EPS (Encapsulated PostScript) is based on PostScript, but was specifically intended to encapsulate graphics. It contained the ability to lock images and layouts, and became somewhat of an industry standard for transferring images to commercial printers, however it has been surpassed by the more versatile PDF format.

The PDF (Portable Document Format) is based on PostScript, and like EPS, contains the ability to lock editing. Unlike EPS however, it contains much more functionality including links, hotspots, metadata, the ability to embed fonts to travel with the document, and accessibility features.

Bibliography

1. About.com, *Graphics Software*, Nov 2007 http://graphicssoft.about.com/library/glossary/bldefmetafile.htm

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

3. Adobe Systems Inc., *Whitepaper-- PDF as a Standard for Archiving,* 2002. http://www.adobe.com/enterprise/pdfs/pdfarchiving.pdf

4. Aitken, Peter. *I Never Metafile I Didn't Like*, 3 November 2006. http://www.devsource.com/article2/0,1895,2050820,00.asp

5. Audio Engineering Society, http://www.aes.org/publications/downloadDocument.cfm?accessID=1470316200012 2117

6. Aware Systems, *The BigTIFF File Format Proposal,* 2007 http://www.awaresystems.be/imaging/tiff/bigtiff.html

7. Bachmann, Erik. *Xbase Data File (*dbf),* 18 Aug 2007 www.clicketyclick.dk/databases/xbase/format/dbf.html#DBF_STRUCT

[3] *File Formats and Compression*, Technical Advisory Service for Images Advice Paper March 2005 www.tasi.ac.uk

8. Biblioscape, *Rich Text Format (RTF) Version 1.5 Specification*, 2007. http://www.biblioscape.com/rtf15_spec.htm#Heading1

9. Broadcastpapers.com, http://www.broadcastpapers.com/sigdis/Snell&WilcoxMXF01.htm

10. Broadcastpapers.com, *The online library for technical & business whitepapers,* November 2007. http://www.broadcastpapers.com/whitepapers.cfm?objid=2

11. 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

12. Brooks, Alfred A. *Overview – ISO/IEC 8211:1994.* 1996

13. Brown, Adrian. *Digital Preservation Guidance Note: 4, Graphics File Formats,* The National Archives, United Kingdom, 9 July 2003. http://www.nationalarchives.gov.uk/documents/graphic_file_formats.pdf

14. 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.

15. Brown, David. *Guidelines for Computer File Types, Interchange Formats and Information Standards*, Library and Archives Canada, Version 1, 28 June 2004.

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

17. Canadian Council on Geomatics, *Standard File Exchange Format for Digital Spatial Data, Version 2.3*, October 1994. www.cits.rncan.gc.ca/fich_ext/1/text/products/ntdb/ccogif.pdf

18. Canadian Standards Association, *Beyond Canada: health informatics around the globe*, 2007. http://www.csa.ca/standards/health_care/newsletter/default.asp?load=news5&language=english

19. CARIS, *Geomatics Software Solutions,* 2007 www.caris.com

20. Carson, Steve. *Basic Image Interchange Format (BIIF),* GSC Associates Inc., 2007.
http://www.acm.org/tsc/biif.html

21. Committee on Earth Observing Satellites (CEOS). *The CEOS Working Group on Information Systems and Services (WGISS),* 2007
http://wgiss.ceos.org/ceos.htm

22. CoOL, a project of the Preservation Department of Stanford University Libraries and Academic Information Resources, Aldus/Microsoft Technical Memorandum, *TIFF Revision 5.0,* November 2007.
http://palimpsest.stanford.edu/bytopic/imaging/std/tiff5.html

23. Cudlip, W. *Guidelines on Standard Formats and Data Description Languages Version 1.0.* Committee on Earth Observation Satellites. 1998.

24. Curtin, Dennis P. *Sensors, Pixels and Image Sizes*, 2007.
http://www.shortcourses.com/pixels/colourdepth.htm

25. Data Compression Dogma, *What is the state of the art in lossless image compression?,* 26 October 2006
http://datacompression.dogma.net/index.php?title=FAQ:What_is_the_state_of_the_art_in_lossless_image_compression%3F#Benchmarks

26. Developer Shed*, Bringing Yourself Up to Speed with AAC, MP3, and Digital Audio,* November 2007.
http://www.devhardware.com/c/a/Software/Bringing-Yourself-Up-to-Speed-with-AAC-MP3-and-Digital-Audio/3/

27. Doughty, Mike. Mike's Sketchpad, *A Little Bit More About Color*, 2007.
http://www.sketchpad.net/basics6.htm

28. Durham University, *Accessibility Glossary,* 2007.
http://www.dur.ac.uk/its/services/web/accessibility/glossary/

29. ER Mapper, *Geospatial Imagery Solutions Forum*, 2007.
http://forum.ermapper.com/viewforum.php?f=11

30. European Broadcast Union (EBU), *Broadcast Wave Format (BWF) User Guide*, 11 May 2007.
http://www.ebu.ch/en/technical/publications/userguides/bwf_user_guide.php

31. Federal 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

32. FileFormatInfo, Encyclopedia of Graphics File Formats, *Microsoft Windows Metafile File Format Summary.*
http://www.fileformat.info/format/wmf/

33. Future Publishing Limited, *Video Codecs,* November 2007.
http://www.pcanswers.co.uk/tutorials/default.asp?pagetypeid=2&articleid=30900&subsectionid=781&subsubsectionid

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

35. Glagola, Michael J. mglagola@cox.net' *Digital Image Characteristics; Understanding Pixels, Dots, Samples & Viewing,* Washington Apple Pi iLife SIG, January 2005
http://www.wap.org/imovie/DigitalImagingPresentation.pdf

36. Grand, Mark. *MIME Overview*, 26 Oct 1993
http://mgrand.home.mindspring.com/mime.html

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

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

39. International Telecommunications Union, *Telecommunication Standardization Sector (ITU-T),* 2007
http://www.itu.int/publications/sector.aspx?lang=en&sector=2

40. 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.

41. International Hydrographic Organization, *IHO Transfer Standard for Digital Hydrographic Data, Publication S-57 Edition 3.1*, Nov 2000
www.iho.shom.fr/publicat/free/files/31Main.pdf

42. International Organization for Standardization, *Coding of Moving Pictures and Audio, MPEG-4 Overview,* March 2002.
http://www.chiariglione.org/mpeg/standards/mpeg-4/mpeg-4.htm

43. International Organization for Standardization. ISO/TC211, *Geographic Information/Geomatics Scope.* (2002).
http://www.isotc211.org/scope.htm#scope

44. International Organization for Standardization, ISO/IEC 8859-1:1998 Information technology -- 8-bit single-byte coded graphic character sets -- Part 1: Latin alphabet No. 1, 2007.
www.iso.org/iso/en/CatalogueDetailPage.CatalogueDetail?CSNUMBER=28245&IC1=35&ICS2=40&ICS3

45. 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

46. Joint Photographic Experts Group, *JPEG 2000 FAQ.,* 2007.
http://www.jpeg.org/.demo/FAQJpeg2k/index.htm

47. Korpela, Jukka. *A tutorial on character code issues*, 13 July 2007.
http://www.cs.tut.fi/~jkorpela/chars.html

48. Kunze, John. California Digital Library, *WARC: an Archiving Format for the Web,* 22 September 2005
http://www.iwaw.net/05/kunze.pdf

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

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

51. Library of Congress, *Digital Preservation, Digital Formats, Sound Quality and Functionality Factors*, 07 March 2007.
http://www.digitalpreservation.gov/formats/content/sound_quality.shtml

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

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

54. Moving Pictures Experts Group. *The MPEG Home Page,* November 2007.
http://www.mpeg.org/

55. Mpeg***TV***, *The reference site for MPEG*, November 2007.
http://www.mpeg.org/MPEG/

56. NCH Swift Sound, *Audio File Formats*, November 2007.
http://www.nch.com.au/acm/formats.html

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

58. NOAA Satellite and Information Service, National Satellite, Data and Information Services (NESDIS), *NOAA Metadata Manager and Repository,* 2007
www.ngdc.noaa.gov/nmmr/public/viewRecord.do?xmlstyle=FGDC&edit=&recuid=1858&recordset=NCDCà

59. OASIS, *Open Document Format for Office Applications (OpenDocument) TC,* 2007.
http://www.oasis-open.org/committees/tc_home.php?wg_abbrev=office

60. Open GIS Consortium Inc. *OpenGIS Geography Markup Language (GML) Implementation Specification*. Document OGC 02-023r4, Version 3.0. editors, Simon Cox, et. al., January 2007.
www.opengis.org/

61. Oracle Corporation, *Image File and Compression Formats,* 2003.
http://www.stanford.edu/dept/itss/docs/oracle/10g/appdev.101/b10829/mm_formats.htm

62. PDF Tools AG, *White Paper – PDF Primer,* 6 October 2005.
http://www.pdf-tools.com/public/downloads/whitepapers/whitepaper-pdfprimer.pdf

63. Pearson, Glenn and Michael Gill, *"An Evaluation of Motion JPEG 2000 for Video Archiving"*, Proc. Archiving 2005 (April 26-29, Washington, D.C.), IS & T (www.imaging.org), pp. 237-243.
http://archive.nlm.nih.gov/pubs/pearson/MJ2_video_archiving.pdf

64. Peterson, Kit A. Digital Conversion Specialist, Prints & Photographs Division, Library of Congress, Washington, D.C. 20540-473, *Introduction to Basic Measures of a Digital Image for Pictorial Collections,* June 2005.
http://www.loc.gov/rr/print/tp/IntroDgtlImage.pdf

65. Radiological Society of North America, *DICOM (Digital Imaging and Communications in Medicine), The Value and Importance of an Imaging Standard,* 2007.
http://www.rsna.org/Technology/DICOM/

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

67. RealNetworks Inc., *RealAudio 10*, 2003.
http://docs.real.com/docs/rn/datasheet/RA_10_Datasheet_S203.pdf

68. RealNetworks Inc., 2007
http://www.realnetworks.com/resources/howto/audio_video/video.html

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

70. Red Hat, Inc., *DocBook: A Tutorial for Hackers and Writers,* 2007.
http://www.linuxshowcase.org/2000/2000papers/papers/mason/mason_html/slide006.html

71. Roelofs, Greg., *An Open, Extensible Image Format with Lossless Compression*, 6 September 2007.
http://www.libpng.org/pub/png/

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

73. Ruth,Mike. *GeoTIFF FAQ Version 2.3*, February, 2005
http://remotesensing.org/geotiff/faq.html#What%20is%20GeoTIFF%20and%20how%20is%20this%20different%20from%20TIFF?

74. TeamCom Books, *The MP3 and Internet Audio Handbook, Your Guide to the Digital Music Revolution!,* March 2000.
http://www.teamcombooks.com/mp3handbook/MP3_Handbook.htm

75. Techsmith,

http://fr.techsmith.com/products/studio/tutorials/1104.asp

76. Unicode Inc., *The Unicode Standard, Version 3.0,* 2007
www.unicode.org/book/u2.html

77. United Kingdom Office for Library and Information Networking (IKOLN). *NOF-Digitization 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).

78. United States Geological Survey, Digital Line Graph Standards, 29 Aug 2007
http://rockyweb.cr.usgs.gov/nmpstds/dlgstds.html

79. United States Geological Survey, Digital Elevation Model Standards, 29 Aug 2007
http://rockyweb.cr.usgs.gov/nmpstds/demstds.html

80. United States Geological Survey*, Spatial Data Transfer Standard,* 2007
http://mcmcweb.er.usgs.gov/sdts

81. United States General Service Administration. *Government Without Boundaries: A Management Approach to Intergovernmental Programs*. Office of Intergovernmental Solutions. May. (2002).

82. United States National Snow and Ice Data Center, *Global Digital Sea Ice Data Bank (GDSIDB),* 2003)
http://nsidc.org/noaa/gdsidb/s3development.html

83. University of Connecticut, *The file types*, 12 November 2007
http://www.gifted.uconn.edu/siegle/HonorsSeminar/filetype.html

84. Usdin, B. Tommie et. al. What is SGML? Mulberry Technologies, Inc. Rockville, MD. 1997.

85. Web3D Consortium, GeoVRML Working Group, *Standard VRML97 1.0 (Virtual Reality Modeling Language),* 2007.
www.ai.sri.com/geovrml/1.0/

86. Web3D Consortium, GeoVRML Working Group, *Standard VRML97 1.1 (Virtual Reality Modeling Language),* 2007.
www.ai.sri.com/geovrml/1.1/

87. Webopedia.com, *CCITT, Comité Consultatif International Téléphonique et Télégraphique,* November 2007.

http://www.webopedia.com/TERM/C/CCITT.html

88. Webopedia.com, *Data Compression,* 2007. http://www.webopedia.com/TERM/d/data_compression.html

89. Wikipedia, The Free Encyclopedia, *Linear Pulse Code Modulation used in communications (or LPCM)*, Nov 2007 http://en.wikipedia.org/w/index.php?title=LPCM&oldid=88811980>.

90. Wikipedia, The Free Encyclopedia, *Sound quality*, 20 Sept 2007 http://en.wikipedia.org/wiki/Sound_quality

91. Wikipedia, The Free Encyclopedia, *Windows Media Audio (WMA)*, 10 November 2007. http://en.wikipedia.org/wiki/Windows_Media_Audio

92. Wikipedia, The Free Encyclopedia, *Ogg Vorbis,* 11 November 2007. http://en.wikipedia.org/wiki/Ogg_Vorbis

93. Wikipedia, The Free Encyclopedia, *Audio File Format,* 8 Nov 2007. http://en.wikipedia.org/wiki/Audio_file_format

94. Wikipedia, The Free Encyclopedia, *Document File Format*, 2007 http://en.wikipedia.org/wiki/Document_file_format

95. Wikipedia, The Free Encyclopedia, *GeoTIFF*, 28 August 2007. http://en.wikipedia.org/wiki/Geotiff

96. Wikipedia, *NetCDF (Network Common Data Form),* November 9, 2007. en.wikipedia.org/wiki/Netcdf

97. Wikipedia, The Free Encyclopedia, *Vector graphics*, 6 November 2007. http://en.wikipedia.org/wiki/Vector_graphics

98. XIPH.ORG, *Vorbis,* November 2007. http://www.vorbis.com/

99. XIPH Open Source Community, *Theora*, November 2007. http://www.theora.org/theorafaq.html#10