Canadian Heritage | Patrimoine canadien

Canadian Digital Cultural Content Initiative (CDCCI)

Standards and Guidelines for Digitization Projects

Version 2.2
Last Updated: August 19 2001

Canada

Acknowledgements

The editors wish to acknowledge the contribution of the Interdepartmental Interoperability Forum (IIF) for their contributions to this document.

This document was written and edited by Mark Alexander and Terry Kuny at the Department of Canadian Heritage. The body of work within the document draws on work accomplished in the United Kingdom on interoperability, by permission of the authors of the original documents.

UK Office for Library and Information Networking (UKOLN). *nof-digitize Technical Standards and Guidelines*, Version Two, Revised November 2000.

URL: http://www.peoplesnetwork.gov.uk/nof

Joint Information Systems Committee (JISC). *Working With the Distributed National Electronic Resource.* February 2001.

URL: http://www.jisc.ac.uk/dner/programmes/guidance/DNERStandards.html

The editors gratefully acknowledge UKOLN and JISC for their generous contributions to the creation of this document.

Updates to this Document

This document will be updated regularly. The latest version is found at:

http://www.pch.gc.ca/cdcci-icccn/eng/pubs.htm

TABLE OF CONTENTS

Introduction to the CDCCI

Background

The Canadian Digital Cultural Content Initiative (CDCCI) is the Department of Canadian Heritage's partnership with private and public institutions and organizations across Canada. The primary objective of the CDCCI is to bring Canadian culture into the digital age.

More specifically, the CDCCI is designed to stimulate the creation and production of Canadian digital cultural content in both official languages and promote a significant, identifiable Canadian presence online that reflects Canada's cultural diversity.

CDCCI funding programs will help to build *meaningful* and s*eamless* access to content that helps deepen an understanding of Canada and of our rich diversity for the benefit of Canadian youth and students.

Fifty (50) percent of CDCCI funds are dedicated to the development of French-language content on the Internet, and priorities are placed on funding projects that focus on youth and collaborative projects that create partnerships from both the public and the private sectors to realize their goals.

Content Integration:
A Gateway to Canadian Cultural Content On-line

The CDCCI intends to integrate access to content digitized through its funding programs on a Canadian cultural portal site to realize its goal of building meaningful and seamless access to digital content. As the portal is at the time of this writing still under development, this document will serve to inform funding recipients of basic principles concerning standards which will: a) ensure that their projects make maximum use of the funding resources received from the CDCCI and b) ensure that the content can be easily integrated into the portal at a later time.

For more information concerning the portal, please contact the Secretariat of the CDCCI directly.

Areas for Consideration

Important areas for consideration for CDCCI funded projects are:

Accessibility: It is important that CDCCI funded resources are as accessible as possible and are made publicly available using open standards and non-proprietary formats. If this material is to be a true national resource it must be accessible to all citizens including those with disabilities.

Interoperability: It is important to be able to share content seamlessly between projects and users, to be able easily to use content without specialist tools, to be able to discover and interact with content in consistent ways, to be able to manage it effectively.

Preservation: It is important to secure the long-term future of materials, so that the benefit of the investment is maximized, and the cultural record is maintained in its historical continuity and media diversity.

Benefits of Interoperability

Seamless Access to Canadian Cultural Resources: A key objective of the CDCCI is to encourage the creation of intellectually coherent digital cultural content which may be easily discovered and accessed by Canadian users wherever they may be -- and to facilitate opportunities for the creation of multiple points of access to content funded by the CDCCI, including the portal site.

Promotion of Best Standards and Practices: The CDCCI will facilitate the development of standards and best practices for the creation of digital Canadian cultural content and will use funding opportunities to promote these as widely as possible. A consensus- based approach to the adoption of standards and best practices has been facilitated through the CDCCI with Canada's key federal cultural institutions who participate in regular discussion forums on this topic. The CDCCI will endeavor to ensure that this document reflects the latest developments within the Canadian cultural community.

Requirements and Guidelines

This document outlines technical requirements that projects must meet and also offers guidance on developing technical issues. Requirements are standards that are widely accepted and already in current use. All projects must implement any standards that are identified as requirements. This document also offers guidance. The standards covered in the guidance sections are still in development. It is probable that these standards will become widely accepted during the lifetime of the programme but they cannot be identified as a requirement with full confidence yet. Projects should demonstrate awareness of these standards and their potential applications.

Vocabulary

The words 'must, should and may' when printed in bold text have precise meanings in the context of this document.

MUST: This word indicates an absolute technical requirement with which all CDCCI projects must comply.

SHOULD: This word indicates that there may be valid reasons not to treat this point of guidance as an absolute requirement, but the full implications should be understood and the case carefully weighed before it is disregarded. 'Should' has been used in conjunction with technical standards that are likely to become widely implemented during the lifetime of the project but currently are still gaining widespread use.

MAY: This word indicates that the topic deserves attention, but projects are not bound by CDCCI advice. 'May' has therefore been used to refer to standards that are currently still being developed.

Content Creation

General Principles

This section covers the actual creation of digital resources by CDCCI funded projects. Technical standards decisions that are made at this stage will have a fundamental impact on the manageability, accessibility, viability and interoperability of the resources created.

The process of deciding what to digitize anticipates all the stages of project management. The use of digital resources not only depends on the nature and importance of the original source materials, but also on the nature and quality of the digitization process itself. Firstly on how well relevant information is captured from the original, and then on how the digital resources are organized, indexed, delivered to users, and maintained over time.

The following criteria should be taken into account in the content selection process:

1. **Relevance to the actual or potential user needs** – materials **MUST** have sufficient intrinsic value to ensure interest. Projects are expected to create unique digital resources or to augment existing collections.

2. **Educational Component** – all CDCCI content created with learning and teaching components **SHOULD** show clear evidence of its relevance to educational needs or learning objectives and demonstrate an appropriate strategy for its effective use for these purposes.

3. **Rights and permissions for electronic distribution MUST** be securable within the parameters of the project and funding available.

4. **Quality of the original source material** – materials to be digitized **SHOULD** be of the highest quality possible, capture from original materials is generally preferable and especially where maximum information capture/fidelity to original sources is of importance.

5. **Metadata** – Resources created **MUST** be supported by the appropriate metadata; such metadata should describe the material comprehensively and coherently as well as providing clear pathways to the user attempting to navigate within the particular information environment.

6. **Currency and timeliness** – some products retain their value only if there is a commitment to providing the most up-to-date information therefore commitments to maintenance and quality of service need to be assured. Plans for long term maintenance of collections **MUST** be demonstrated within the parameters of all projects funded by the CDCCI.

7. **Coverage of the collection** – the new resource **MUST** either represent a new collection or be a worthwhile addition to existing collection(s) of resources. The production of duplicated material **MUST** be avoided.

Preparation

Preparation is one of the most neglected areas of the digitization chain, but if done well can help avoid many of the most common errors or quality defects that occur in large-scale digitization projects.

Preparation falls into two categories: preparation of originals, and preparation of the capture process (including metadata). Where digitization is outsourced, this work may be divided into, that done by the owner of originals before the materials are moved to the digitization site, and that done by the operator of digitization services before any capture takes place. If digitization is taking place in house, both processes may be carried out by the same group. However it is important to build in the time to both prepare originals as well as set up the capture process before commencing.

Further Information

TASI Framework. *Image handling and preparation.*

URL: http://www.tasi.ac.uk/framework/capture/imagehand.html

Preparation for capture of digital information.

URL: http://heds.herts.ac.uk/resources/papers/jidi_fs.pdf

Digitization: Strategic and Management Issues.

URL: http://heds.herts.ac.uk/resources/Papers/HEDSITForum.pdf

Digitization how much does it really cost?

URL: http://heds.herts.ac.uk

Production Equipment and Facilities

The CDCCI programme provides only minimal resources for the development of digitization infrastructure. Organizations are expected to use industry standard and appropriate equipment and techniques for the capture and storage of digital resources prepared for CDCCI purposes. It is important to consider whether the resource should be digitized in house or whether this work should be out-sourced. Several factors govern this decision-making process including, cost, viability of movement of originals, skills set available in house etc. Applicants will be required to supply justifications for decisions about digitization, for example, to demonstrate that lowest cost bids were solicited for scanning or data conversions, or to justify the purchase of equipment. Information to assist decision-making can be found below.

Further Information

HEDS. *The Digitization Process: Set up an in-house scanning unit or use a bureau?*

URL: http://www.ukoln.ac.uk/nof/support/help/papers/digitization.htm

JIDI (JISC Image Digitization Initiative) Digitization Feasibility Study.

URL: http://heds.herts.ac.uk/resources/papers/jidi_fs.pdf

VADS/TASI. *Guide to Good Practice, Section 3.*

3.1 Introducing digital images and image creation

3.2 Digital images

3.3 Selecting image file formats

3.4 Image manipulation and imaging software

3.5 Image capture equipment: scanners

3.6 Image capture equipment: digital cameras

URL: http://vads.ahds.ac.uk/guides/creating_guide/contents.html

TASI Framework.

URL: http://www.tasi.ac.uk/framework/capture/hwandsw.html

RLG/DLF Guidelines. *Guides to Quality in Digital Resource Imaging; Selecting a Scanner.*

URL: http://www.rlg.org/visguides/visguide2.html

Overview of Scanners.

URL: http://www.gii.getty.edu/intro_imaging/11-Scan.html

How digital cameras work.

URL: http://www.cmospro.com/howdcw.htm

Workflow (or procedures management)

The CDCCI programme expects projects to establish sound workflow processes that provide for the most efficient capture and management of digital resources.

Workflow for digital content capture is the process of establishing and developing the chain of events that lead from an analogue work to a digital one. This may involve locating, un-shelving and preparing originals for digitization, capturing work, creating metadata, post processing and storage. This work will probably be done by a team of people rather than one individual, often third parties, for example with digital imaging, a photographer or digitization bureau may be an intrinsic part of this chain. The relationship between these parties, and how information, as well as digital or original objects are transferred defines the pattern of workflow. Smooth workflow where effort is not duplicated and which proceeds logically from one stage to the next is crucial.

For CDCCI funded projects, procedures **SHOULD** be put in place for managing the digitization chain effectively and to keep track of the originals and digital surrogates as they are created. This process of tracking workflow can be time consuming. However it is an essential part of the process and should not be overlooked to the detriment of the end product but also potentially to the safety and condition of the original objects.

Further Information

JIDI/TASI. *Workflow Guidelines.*

URL: http://www.ilrt.bris.ac.uk/jidi/workflow.html

VADS/TASI. *Guide to Good Practice.* (some dedicated sections on workflow.)

URL: http://vads.ahds.ac.uk/guides/creating_guide/sect52.html

URL: http://vads.ahds.ac.uk/guides/creating_guide/appendix2.html

Capture Standards

Background

The following section establishes some baseline capture standards for projects developing digital content to be funded by the CDCCI. The following points should be adhered to during content capture:

- Standard formats **MUST** be used wherever possible. They are more likely to have a good shelf life making the content produced suitable for archiving and preservation. It is recognized that some areas, particularly time-based media do not have stable nonproprietary formats currently available for capture.

- One purpose of digitization is to provide an archival copy that will mitigate the possibility of rescanning being necessary in the future. Therefore scanning at highest possible resolution (300-600 dpi depending upon the content) **SHOULD** be achieved.

- Appropriate web level surrogates for use for on-line delivery **MUST** be generated where this is relevant to a project's dissemination goals. For example, high-quality digital images appropriate for preservation are large and unwieldy to use over a computer network, so projects **MUST** consider the development of specifications for access-quality images. Lower resolution images - whose digital files should be smaller (in bytes) - can be more easily handled by personal or institutional computer systems. Such images would be lower in either spatial resolution ("dots per inch") or tonal resolution ("bits per pixel") or both, and **SHOULD** be derived from the highest resolution images available.

- Overall the capture format chosen **MUST** be appropriate for the content and the mode of content delivery.

Capture Guidelines

Digital imaging techniques offer the opportunity to easily and quickly capture images or other archival materials. These can act as a means of improving access to materials or as a step toward preservation or both. In either case, projects developing digital content **SHOULD** refer to a set of practical guidelines that assist them in making decisions about image quality, presentation/display options, and storage formats.

Further Information

HEDS. *Digitization Feasibility Studies.*

URL: http://heds.herts.ac.uk/resources/papers/jidi_fs.pdf

URL: http://heds.herts.ac.uk/resources/papersF.html

TASI Framework. *Building Image Archives.*

URL: http://www.tasi.ac.uk/framework/framework.html

Research Libraries Group & National Preservation Office (UK). *Guidelines for Image Capture.*

URL: http://www.rlg.ac.uk/preserv/joint

Research Libraries Group. *Digital Imaging for Photographic Collections: foundations for technical standards.*

URL: http://www.rlg.org/preserv/diginews/diginews3.html#com

Kenny A and Chapman S. (1996). *Digital Imaging for libraries and archives.* Cornell University.

Creating Digital Resources for the Visual Arts; VADS, Section 3.

URL: http://vads.ahds.ac.uk/guides/creating_guide/contents.html

Research Libraries Group. Digital Library Federation. *Guides to Quality in Visual Resource Imaging.*

URL: http://www.rlg.org/visguides/

Normes, techniques et standards de qualité recommandés par le réseau Info-Muse de la SMQ. Projet de numérisation du Réseau Info-Muse. L'Observatoire SMQ (Société des musées québecois)

URL: http://www.smq.qc.ca/publicsspec/smq/activites/numerisation/normes.phtml

Numérisation du patrimoine culturel: informations techniques. Le Ministère de la culture et de la communication.

URL: http://www.culture.fr/culture/mrt/numerisation/fr/technique/technique.htm

Still Images

- Images **MUST** be created and stored using the TIFF or PNG format for archival purposes. These formats are non-compressed, and non-lossy and are therefore suitable for initial capture purposes.

- Images **MUST** be provided on the Web as GIF (for line-drawings) or JPEG (for photographs) formats. PNG MAY be used as an alternative format to GIF.

- Images **SHOULD** be available in a variety of sizes for networked delivery including a thumbnail size.

Further Information

JPEG

URL: Specification: http://www.w3.org/Graphics/JPEG/itu-t81.pdf

Further information : http://www.w3.org/Graphics/JPEG/

PNG
URL: Specification: http://www.w3.org/TR/REC-png.html Further information : http://www.libpng.org/pub/png/

Moving Images

Methods and techniques of audio-video compression are evolving rapidly. Currently, lossless audio-video compression cannot be delivered over public networks such as the Internet. Therefore, there must be compromises made between encoding for distribution and encoding for preservation.

Moving Picture content **SHOULD** be encoded using either Apple Quicktime, Microsoft Windows Media Player Real Video, MPEG2 or MPEG4. MPEG1 SHOULD not be used except where a need can be demonstrated to the CDCCI. Where codecs are not implicit, encoding should be done using codecs supplied free with the appropriate players and which enable use on a wide variety of platforms.

Particular attention should be paid to data rates used during compression. As Internet audiences access the networks via a mix of modems, cable modems, DSL, satellite, and higher bandwidth connections, the desired audience might be comprised of a mix of low to high speed access. Therefore, the encoding process might create different files for the different bandwidth profiles. Some systems include those different data rates in the same file (Real's SureStream, Windows Intelligent Stream). Appropriate mention of the data rates comprised within the same file or several files **SHOULD** be described.

- Formats for re-encoding: DV, miniDV, DVCam, DVCPro, DVCPro 50, digiBeta and BetaSP are all acceptable.

Sound

As with video, there may be a need to encode sound on different profiles for distribution and for preservation. Multiple bit rates files are also possible, so attention **SHOULD** be given during the encoding process to accommodate the different target audiences.

Sound content **SHOULD** be encoded using either MPEG1 Audio Layer 3, WAV files, Apple Quicktime, Microsoft Windows Media Player or RealAudio. Where codecs are not implicit, encoding **SHOULD** be done using codecs supplied free with the appropriate players and which enable use on a wide variety of platforms.

Sound formats for re-encoding: CD-Audio (44Khz @ 16Bits) or DAT (44Khz @ 16Bits or above) are recommended.

Further Information

MPEG - Moving Pictures Expert Group (Working Group of ISO/IEC).
URL : http://www.cselt.it/mpeg/

Overview of the MPEG 7 standard (Multimedia Content Description).
URL : http://www.cselt.it/mpeg/standards/mpeg-7/mpeg-7.htm

BBC - Standard Media Exchange Framework (SMEF™) Data Model.
URL : http://www.bbc.co.uk/guidelines/smef/

JISC Advisory Service for Moving Images and Sound.
URL: http://www.bufvc.ac.uk/

Guidance on encoding can be found at:
URL: http://www.codeccentral.com/

Apple's QuickTime Developper Ressources
URL: http://developer.apple.com/quicktime/

Microsoft Windows Media Technologies
URL : http://www.microsoft.com/windows/windowsmedia/en/default.asp

RealNetworks devzone tutorials
URL: http://www.realnetworks.com/devzone/tutorials/index.html

PADS. *Guide to Good Practice - Creating Digital Performance Resources.*
URL: http://www.pads.ahds.ac.uk/padsGGPPerformance

Text (e-books, journals, other texts)

- Where possible, text based content **SHOULD** be created and managed in a structured format suitable for delivery as HTML. In most cases storing text-based content as HTML, XHTML, XML or ASCII text will be the most appropriate option.

- Where HTML is chosen as the storage format, the latest version recommended by the World Wide Web Consortium (W3C) **MUST** be used. HTML**Error! Bookmark not defined.** and XML documents and CSS stylesheets **MUST** be validated using services such as those freely supplied by the W3C and other organizations.

- Text-based content **SHOULD** be delivered as HTML or XHTML in most cases, though the use of other XML DTDs may sometimes be appropriate. Where HTML is chosen as the delivery format, the latest version recommended by W3C (currently HTML 4) or the latest version of XHTML (currently 1.0) **MUST** be used. HTML and XML documents **MUST** validated against a published DTD and/or using services such as those freely supplied by the W3C and other organizations.

- In some cases, delivery in formats such as PDF, RTF or ASCII **MAY** be appropriate.

-

Further Information

Oxford Text Archive. *Creating and Documenting Electronic Texts.*
URL: http://ota.ahds.ac.uk/documents/creating/

HTML
URL: Specification: http://www.w3.org/TR/html4/
Further information : http://www.stars.com/Authoring/HTML/

XHTML
URL: http://www.w3.org/TR/xhtml1/
Further information : http://www.w3.org/MarkUp/Activity

CSS
URL: http://www.w3.org/TR/REC-CSS2/
Further information : http://www.w3.org/Style/CSS/

XML
URL: http://www.w3.org/TR/2000/REC-xml-20001006
Further information : http://www.xml.com/pub/

HTML Validation Service.
URL: http://validator.w3.org/

CSS Validator.
URL: http://jigsaw.w3.org/css-validator/

HTML Tidy.

URL: http://www.w3.org/People/Raggett/tidy/

Access Adobe.

URL: http://access.adobe.com/

Cataloguing and Metadata Standards

Background

In managing digital collections and supporting their use, the need for accurate information or metadata (structured data about data) is essential if we are to understand the identity, context and usefulness of a collection and locate it in a networked environment.

The need to improve the effectiveness of searching for information resources on the Internet has prompted the development of simplified metadata standards that can be used at relatively low cost. (The term "metadata" refers to data that relates to one or more information resources, supporting their discovery or management).

These metadata standards are intended to improve the general efficiency of searching for information on the Internet. However it is important to note that they do not supplant the need for good and consistent cataloguing of digital collections. The creation of standards compliant metadata is crucial in order to provide usable and sustainable resources.

Good Practice for Cataloguing and Metadata Creation

For the purposes of this document, cataloguing and metadata creation refers to the process of creating structured descriptions that provide information about any aspect of a digital resource.

This information may include technical information about both the digital entity, *e.g.* a scanned image, but also the photograph or original object from which the digital entity is derived.

This information is also likely to describe the whole of a resource, *e.g.* the whole digital or physical collection as well as its component parts.

Other metadata that may be relevant is that which describes the process of digitization, which documents the decisions that have been made during the resource creation process, and provides a comprehensive and authoritative way of referring to the finished product and its contents. (Types of specific process metadata may be administrative metadata, technical metadata and preservation metadata)

There are some important general points to bear in mind when creating metadata and cataloguing digital resources for CDCCI funded projects:

- CDCCI funding recipients **MUST** provide metadata about their projects to CDCCI in both French and English of equal quality using RSLP templates. This information will be used for the development of Canadian "collection finder" or "yellow pages" services. Information about RSLP can be found in the section "Collection Description Standards" below.

- All projects **SHOULD** attempt to make all resources accessible through the use of both French and English metadata where possible.

- CDCCI funding recipients **SHOULD** use an existing standard for digital resource description and use recognized standards for terminology and relevant controlled vocabularies. Examples can be found in the section "Terminology Standards and Controlled Vocabularies" below.

- Metadata **SHOULD** always be internally consistent, *e.g.* consistent spelling, use of classification schemes or terminology.

- Pay equal attention to creating all the types of metadata described above, *e.g.* technical metadata is needed as much as metadata describing the subject or content of a resource. With the passage of time without the metadata that tells us about the digital object we will may not be able to identify the format of a resource.

- CDCCI funding recipients **MUST** provide guidelines detailing the cataloguing and metadata creation for their project. These should explain any rules of how fields should be populated or how different categories of information should be created.

- To make sure quality is maintained cataloguers **SHOULD** be trained before being permitted to create metadata.

- CDCCI funding recipients **MUST NOT** create their own standards for metadata, or substantially amend or add to an existing standard for internal use without providing an adequate rationale for this process or documenting what changes have been made.

Forum for Metadata Schema Implementers.

URL: http://www.schemas-forum.org/

VADS. Guide to Good Practice. *Standards for Data Documentation.*

URL: http://vads.ahds.ac.uk/guides/creating_guide/sect41.html

Open Archive Initiative.

URL: http://www.openarchives.org/

Digital Resource Description Standards

The provision of descriptions of collections and the objects contained within them is essential for the access to and maintenance of collections and objects.

It is also important to note that the information architecture of Canadian Heritage and Government Online services is still evolving. In the next months it is expected that more detailed definitions of metadata standards will emerge to support interoperability in the longer term. Use of the following standards will provide a useful bridge to any future resource discovery services that are developed within government and within various content communities.

Dublin Core

Prominent amongst general standards for the creation of metadata is the Dublin Core metadata set. Dublin Core is being developed as a generic metadata standard for use by libraries, museums, archives, government and other publishers of online information. This standard may be applied broadly to citation and full text descriptions, and may support interoperability between a range of other cataloging or metadata schemes. Dublin Core is being adopted by the Government of Canada for cross-searching of government resources.

- All CDCCI projects **MUST** either use the Dublin Core as a basis for their metadata standard **OR** provide a mapping to Dublin Core metadata that explains the relationship between the metadata standards they are using and the Dublin Core.

- All projects that provide database access to item level description **MUST** be able to provide a set of Dublin Core metadata for the purposes of harvesting by cross-collection search tools.

Further Information

Dublin Core Metadata Initiative. This is the authoritative reference for this standard.

URL: http://dublincore.org/

Using Dublin Core.

URL: http://dublincore.org/documents/usageguide/

Paul Miller. *Metadata for the Masses.* A useful primer to metadata and Dublin Core. Technical details are dated and the Dublin Core Metadata Initiative site should be referred to for current information.

URL: http://www.ariadne.ac.uk/issue5/metadata-masses/

Collection Description Standards

A variety of tools and services will evolve to facilitate identification of and access to Canadian content projects. To facilitate this "yellow pages" discovery of projects, CDCCI requires that each project provide to CDCCI French and English metadata records for each funded project(s) and collection(s).

- Projects **MUST** provide these records to CDCCI using the RSLP Collection Description schema.

- Projects **MUST** provide these records to CDCCI as an electronic file using the RDF provided by the RSLP tool.

Further Information

RSLP Collection Description.

URL: http://www.ukoln.ac.uk/metadata/rslp/

RSLP Collection Description Tool.

URL: http://www.ukoln.ac.uk/metadata/rslp/tool/

File-naming & Persistent Naming

All projects **MUST** adopt a file-naming convention for digital objects. A well-considered set of file-naming practices will enable portability of objects and sharing of information in a variety of retrieval systems. A file-naming convention helps to manage digitization workflow, assure fewer interoperability issues with any future projects involving aggregation of discrete objects, and eases maintenance for distributed and/or partner projects.

Where possible, a consistent file-naming convention can assist in retrieval and access by users by providing some "intelligence" about where an object comes from. Some digital objects may be reused in places outside of the host institution and intelligent file-naming can assist users in knowing where the object originated or to ensure that objects are not overwritten where they are used for aggregate purposes (for example, collections of thumbnail images).

It is also expected that imposing consistent file-naming will ease the migration to future logical naming schemes such as URNs or DOIs being examined in the Internet community.

CDCCI funded projects **SHOULD** also ensure that objects can be identified with a persistent URL for the purposes of citation, cross-linking and integrated access. Objects retrieved from a database should not have dynamically assigned identifiers (for example, session keys) embedded within the URL as this defeats persistence.

Community Specific Standards

CDCCI promotion of Dublin Core *does not* preclude any project from using rich resource description standards such as MARC, Encoded Archival Descriptions, GIS metadata, or other community specific standards. However CDCCI *does* require that projects to supply, as an output format, a simple Dublin Core record that can be used for cross-searching and harvesting across collections.

While it is important to use or map to Dublin Core for the above purposes, it is probable that you will need to consider using standards that have specific relevance to the community you represent and the systems you already have in place, *e.g.* archival, museum, educational, *etc.*

- Projects **SHOULD** take account of relevant standards specific to their areas when creating metadata.

Learning Materials

Industry Canada, through the Multimedia Learning Group (MLG), is a member of the IMS Global Learning Consortium Inc. IMS is an international organization developing and promoting open specifications for online distributed learning activities such as locating and using educational content, and facilitating interoperability between educational administrative systems.

MLG participates on the IMS Technical Board, co-chairs the IMS Working Group on Learning Design, and is represented on the Accessibility Working Group. The Working Groups develop specifications that are ultimately reviewed and approved by the IMS Technical Board.

Given the development of metadata-based e-learning systems within the federal government, and the growing interest in metadata and e-learning specifications and standards (recommendations concerning e-learning interoperability were made by the Advisory Committee for Online Learning), projects **MAY** be required to provide IMS metadata about their projects for inclusion in appropriate educational resource directories.

Projects **SHOULD** apply the current standards from IMS, CanCore or other educational metadata as they become approved by the government of Canada or other appropriate bodies.

Further Information

Advisory Committee for Online Learning.

URL: http://www.schoolnet.ca/mlg/sites/acol-ccael/

CANCORE

URL: http://www.cancore.ca/

IMS Global Learning Consortium Inc.

URL: http://www.imsproject.org

Terminology Standards and Controlled Vocabularies

Terminology standards and controlled vocabularies are a way of controlling the terms we apply during the process of metadata creation. Their application allows us to ensure that terminology is both internally consistent, e.g. consistent within a given set of metadata but also that it is consistent with the terminology and vocabulary used by others who have also adopted common standards in this manner.

- Where possible, and where these are available, projects **SHOULD** use a relevant terminology standard or controlled vocabulary.

- All federal projects **MUST** use the Treasury Board guidelines and thesauri for applicable subject description.

Further Information

Canadian Subject Headings.

URL: http://www.nlc-bnc.ca/6/23/index-e.html (English)

URL: http://www.nlc-bnc.ca/6/23/index-f.html (Français)

Depository Services Program Thesaurus.

URL: http://dsp-psd.pwgsc.gc.ca/Thesaurus/index-e.html (English)

URL: http://dsp-psd.pwgsc.gc.ca/Thesaurus/index-f.html (Français)

Répertoire de vedettes-matière

URL: www.nlc-bnc.ca/rvmweb/index-e.htm (English)

URL: www.nlc-bnc.ca/rvmweb/index-e.htm (Français)

Statistics Canada

URL: http://www.statcan:80/english/search/thesaurus.htm (English)

URL: http://www.statcan:80/francais/search/thesaurus.htm (Français)

TGN: Thesaurus of Geographic Names.

URL: http://shiva.pub.getty.edu/tgn_browser/

AAT: Art and Architecture Thesaurus.

URL: http://shiva.pub.getty.edu/aat_browser/

wordHOARD: An index to terminology projects and resources relevant to the Museums Sector.

URL: http://www.mdocassn.demon.co.uk/wrdhrd1.htm

MESH: Medical Subject Headings from the National Library of Medicine.

URL: http://www.nlm.nih.gov/mesh/

Database Structures: Selection and Implications

Over time, CDCCI funded projects **MAY** be required to integrate with other similar projects. To facilitate future integration and interoperability of content, it is important that projects document their work and data structures for the future. Following the practices described below will make the process of integration and interoperability work more efficiently.

Designing a suitable data structure is the key to ensuring that the creation of metadata operates smoothly. Creating a data structure is the stage where relationships between different aspects of metadata, *e.g.* object and digital surrogate, technical metadata and content based metadata are defined, as well as the relationship between the intellectual concepts described within the metadata. In a database these relationships are often separated into tables within which information about a single entity or concept is focused. The following considerations should be borne in mind when creating a data structure.

Projects **SHOULD:**

1. Reduce complexity if possible. In general, complex and multi-layered data structures should be avoided. For example a database that has several tables, with different hierarchies of entities within it and different relationships between entities (for example one to many) is inherently more difficult to manage. This is particularly true given that the need will arise to migrate the database through changes in software, platform or staff.

2. Do not create extraneous metadata that you do not need. Consider carefully the intended purpose for the information you are including.

3. Use a database that supports open standards. Select a database that allows you to import and export the metadata into a range of commonly used formats that are independent of a particular platform rather than proprietary. ASCII text based formats are the safest to use for the purpose of exporting and importing data, and your database software should support their use. For example CSV, Comma Separated Variable Format, is a commonly used text format for export of data from a database. ASCII based formats are also more preservation friendly.

4. Document your database design and structure. As above make sure you store documentation about your database design and its rationale, this will allow others to make sense of it.

5. Spend time and attention over the design of the database or other structure that holds the cataloguing information and metadata. Straightforward database structures are generally preferable.

6. Data structures for database projects **MUST** be thoroughly documented in order that you and other third parties can understand your rationale should it become necessary to move or replicate your database for disaster recovery, preservation or other purposes. For example, if you are using a database it is important to document how the fields you have selected are intended to function, and how the tables relate to each other.

Project Web Site Guidelines

Accessibility

- All projects **SHOULD** conform to the W3C Level 1 and Level 2 Accessibility requirements. Any project seeking funding from the CDCCI **MUST** provide a comprehensive explanation of project accessibility planning to the CDCCI. Links to resources on accessibility are listed below under *Further Information on Accessibility Guidelines*. All questions regarding these guidelines should be referred directly to the CDCCI.

General

- All CDCCI funded content websites **MUST** conform with current W3C web standards.

- All project websites **MUST** be valid according to the latest W3C HTML (currently HTML 4.0) or XHTML specifications.

- All project websites using Cascading Style Sheets **MUST** validate using W3C or other third-party validation tools.

- All projects involving federal government agencies **SHOULD** conform with appropriate Treasury Board guidelines for projects.

Further Information on Accessibility Guidelines

Treasury Board of Canada. "Common Look and Feel for the Internet"

URL: http://www.cio-dpi.gc.ca/clf-upe/index_e.asp

Treasury Board of Canada. "Common Look and Feel Self-Assessment Guide"

URL: http://www.cio-dpi.gc.ca/clf-upe/guide/guide_e.asp

W3C Web Accessibility Initiative

URL: http://www.w3.org/WAI/

Web Accessibility Initiative (WAI) Specification.

URL: http://www.w3.org/TR/WAI-WEBCONTENT/

Further Information on Website Resources

World Wide Web Consortium.

URL: http://www.w3c.org

HTML Validation Service.

URL: http://validator.w3.org/

CSS Validator.

URL: http://jigsaw.w3.org/css-validator/

HTML Tidy.

URL: http://www.w3.org/People/Raggett/tidy/

ECMAScript.

URL: http://www.ecma.ch/ecma1/stand/ecma-262.htm

Web Site Auditing and Evaluation

All CDCCI funded projects have accountability requirements that will be met by regular reporting and evaluation. Site visits, web site audits and logfile analysis will be done on CDCCI projects for evaluation purposes. To facilitate this work:

- CDCCI funding recipients **MUST** retain web server logs specific to the CDCCI funded projects for the duration of the project and for a period of two (2) years after rollout of the project. The CDCCI will specify a standard format for reporting aggregated statistics at the time of evaluation For database-driven projects,

usage statistics may also be required. These files will not be released or made available to the public and will be used for project evaluation only.

Programming and Scripting Languages

A number of approaches to server-side scripting may be used, such as using languages such as Perl to provide CGI (Common Gateway Interface) services, server scripting environment such as PHP or ASP, the Java language (and other related components such as JavaBeans), database integration tools or use of content management systems (such as Zope, ColdFusion, etc.) which may have their own scripting environment.

The preferred language for client-side scripting is ECMAscript, the standardised version of JavaScript.

- It should be noted that users may switch off support for ECMAscript or it may not be supported on platforms such as PDAs, digital TV, *etc.* Therefore project web sites **SHOULD** be usable if ECMAscript or programming is disabled or not available.

- ECMAscript **SHOULD** use the DOM (Document Object Model) to manipulate HTML and CSS elements.

- Projects **MAY** use cookies to maintain state information, but since users may switch off support for cookies and cookies may not be supported on platforms such as PDAs, digital TV, etc. project Web sites **SHOULD** be usable if support for cookies is disabled or not available.

Preservation & Records Management

Background

CDCCI recognizes the many difficulties inherent with the long-term preservation of digital information. The following are strategies and approaches that can help to ensure that Canadian digital resources do not become inaccessible prematurely.

CDCCI encourages projects that are concerned about the long-term preservation of digital materials to contact the CDCCI directly for assistance and / or to arrange for deposits of collections for the purposes of storing copies of the material for future authentication, backup and archival purposes with one of Canada's federal cultural institutions.

Key Contacts

National Archives of Canada.

URL: http://www.archives.ca/

National Library of Canada.

URL: http://www.nlc-bnc.ca/media

- Keep store and access areas free of smoke, dust, dirt and other contaminants.
- Store magnetic media away from strong magnetic fields.
- Transport magnetic media in enclosures with space clearances of 50mm.
- Store in a cool, dry, stable and secure environment.
- Acclimatize media before use.
- Use high quality media and devices.
- Keep access devices well maintained and clean.
- Do not place labels on optical disks and/or mark using a pen or pencil. Follow manufacturers recommendations for labeling.
- Minimize handling and use of archival media and/or record number of accesses/use and implement appropriate refreshing.

File Formats

- Use 'open' nonproprietary, well-documented file formats wherever possible.
- Alternatively utilize file formats that are well-developed, have been widely adopted and are 'de facto' standards in the marketplace.
- Identify formats acceptable for the purposes of transfer, storage and distribution to users (these may be distinct).
- Minimize the number of file formats to be managed as far as is feasible/desirable.
- Do not use encryption or compression for archival files if possible.

Media refreshing and reformatting

- Refresh or transfer archive copies to new media at specified times. This should take place:
  - within the minimum time specified by the supplier for the media's viability under prevailing environmental conditions;
  - when new storage devices are installed;
  - when an audit discloses significant temporary or read "errors" in a data resource.
- Employ quality control procedure such as bit/byte or other checksum comparisons with originals to ensure the authenticity and integrity of items after media refreshing.
- Write archive copies with different software to protect data against corruption from malfunctioning or virus- or bug-ridden software.

- Write archive to comparable magnetic media purchased from different suppliers to guard against faults introduced by the media's suppliers into their products or into batches of their products.
- Document actions taken when data resources are copied.
- Retain copies of the digital resource in its original format whenever some information or presentation of the resource may be lost or modified in re-formatting.

Backups and Disaster Recovery Planning

- Create backup copies of data resources at the time of their creation and institute regular backup sessions of data (this may be done automatically over the network).
- Store backup copies on industry standard digital tape or on other approved contemporary media.
- Store backup copies on- and off site. Off-site copies should be stored at a safe distance from on-site copies to ensure they are unaffected by any natural or man-made disaster affecting the on-site copies.

Environmental conditions

- Follow relevant guidance on environmental conditions for storage media.
- Prohibit smoking and eating in the storage and work areas.
- Store away from direct sunlight.
- Provide additional protection in the form of enclosures for media.
- Provide storage facilities that minimize the threat from natural disasters such as fire and flood or to magnetic storage media from magnetic fields.
- Ensure any non-digital accompanying materials (e.g. codebooks, operating instructions) are also stored in appropriate environmental conditions.
- Establish guidance and procedures for acclimatizing magnetic tape if moving between significant variations in temperature (e.g. tapes moving from very cold external conditions should not be used before being acclimatized to warmer internal conditions).

Audit

- Check media periodically for their readability. Such checking may be conducted automatically in mass storage systems according to parameters set by system operators.
- Check the integrity of data files periodically using checksum procedures. Such procedures may be implemented automatically in mass storage systems according to parameters set by system operators.

- Employ appropriate security systems and procedures to protect the authenticity of items in your holdings

Security

- Establish disaster recovery plan.
- Control access to storage facilities and processing areas. Store in separate, preferably lockable area.
- Use passwords and user ids, and other network security procedures.
- Define system and area access privileges for staff.
- Assign specific staff responsibilities for data security and storage facilities.

Further Information

Beagrie, N. and Jones, M. (forthcoming.) *Preservation Management of Digital Materials Workbook*.
URL: http://www.jisc.ac.uk/dner/preservation/workbook/

AHDS. *Guides to Good Practice.*
URL: http://ahds.ac.uk/public/guides.html

Beagrie, N. & Greenstein, D. (1998). *A Strategic Policy Framework for Creating and Preserving Digital Collections.* Version 4.0 (Final Draft). ELib Supporting Study p3. Library Information and Technology Centre, London.
URL: http://ahds.ac.uk/manage/framework.htm

DLM Forum. *Guidelines on Best Practice for Using Electronic Information.*.
URL: http://ispo.cec.be/dlm/documents/guidelines.html

Feeney, M (ed). (1999). *Digital Culture: Maximising the Nation's Investment.* London: The National Preservation Office.
URL: http://www.ukoln.ac.uk/services/elib/papers/other/jisc-npo-dig/intro.html

Lawrence, G.W et al (2000). *Risk Management of Digital Information: A File Format Investigation.*
Council on Library and Information Resources. June 2000. (ISBN 1-887334-78-5).
URL: http://www.clir.org.pubs/reports/reports.html

NOF-digitize Technical Standards and Guidelines. Version One. June 2000.
URL: http://www.peoplesnetwork.gov.uk/nof/technicalstandards.html

Preserving Access to Digital Information (PADI). 'Storage'.
URL: http://www.nla.gov.au/padi/topics/10.html

'Digital Preservation Strategies'.

URL: http://www.nla.gov.au/padi/topics/18.html

TASI. *Recommendations for Digital Preservation.*

URL: http://www.tasi.ac.uk/building/digprint.html

CAMiLEON (Creative Archiving at Michigan and Leeds; Emulating the Old and the New) Project.

URL: http://www.www.leeds.ac.uk/cedars/

Rothenberg, Jeff. (1999). *Avoiding Technological Quicksand: Finding a Viable Technical Foundation for Digital Preservation.* Council on Library and Information Resources. January 1999. (ISBN 1-887334-63-7).

URL: http://www.clir.org/pubs/abstract/pub77.html

INDEX

S

T

V

W

X

Z