Technical Standards and Guidelines for CCO-Funded Initiatives

Canadian Culture Online
25 Eddy St.
Gatineau, Quebec
K1A 0M5
1(866) 900-0001
ccop-pcce@pch.gc.ca

Version 3.0
Last Updated: September 16, 2004

Canada

Updates to this Document

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

http://www.pch.gc.ca/progs/pcce-ccop/pubs/pubs_e.cfm

Table of Contents

Introduction to CCO Standards

Background

This document outlines the technical requirements that Canadian Culture Online (CCO) funded projects must meet. The requirements are based on widely-accepted standards developed by such organizations as the World Wide Web Consortium[1] (W3C) and the International Standards Organization[2] (ISO). All CCO-funded projects must implement the standards identified as requirements.

The standards and guidelines presented in this document should be interpreted as a base line from which institutions can develop additional requirements.

In addition, CCO recommends, but does not require, that recipients apply additional standards in some cases.

Principles and Goals of Technical Standards Guidelines

CCO technical standards are based on the following principles:

- *Accessibility*: Content created, digitized and developed through CCO-funded projects must be made available to all Canadians regardless of geographic or technical boundaries, or disabilities.
- *Visibility:* Standards help to maximize the visibility of CCO-funded projects by facilitating discovery on Internet search engines as well as by other cultural gateway sites.
- *Preservation*: CCO encourages preservation of the digital material it helps create. Long-term preservation strategies maximize the investments made today and maintain a record of the evolution of our online culture.
- *Interoperability*: Well-defined standards help to ensure that content can be shared and operate in conjunction with other content.

1 W3C Web site: http://www.w3c.org
2 ISO Web site: http://www.iso.org

Summary of Technical Requirements

Details on all of the requirements below, as well as best practices for complying with the requirements, are found in the main body of this document (starting on page 6).

Fund recipients are required to:

- conform to accessibility guidelines as defined by the W3C's Web Accessibility Initiative (WAI) Guidelines 1.0. (http://www.w3c.org/TR/WCAG10/)
- develop and document file- and directory-naming rules and guidance for their respective CCO-funded projects
- display text-based content as HTML or XHTML. The latest version recommended by W3C (currently HTML 4.01) or the latest version of XHTML (currently 1.0) must be used
- make text-based content prepared in other formats such as Adobe's PDF available in an alternate format, such as HTML, which is accessible to individuals with disabilities
- use one of the following formats when using multimedia applications:
    - JPEG (photos) or GIF (computer graphics) for graphics on the Web
    - TIFF or PNG for storage of digitized images
      (images can be provided in a second format beyond those mentioned above for the purpose of allowing downloading of images)
    - video, animation and sound file formats that can be played on freely-available plug-ins or players such as QuickTime, RealPlayer, Media Player or ShockWave
- prepare alternate low-bandwidth versions when video and sound files are prepared for delivery in a high-bandwidth environment
- provide alternate, text-based summaries of video, audio and animation files that allow any user to obtain the essence of individual files
- use a database that supports open standards such as SQL and XML
- document their database designs and structures
- use client-side scripting languages that comply with the ECMAScript standard such as JavaScript version 1.5 or JScript version 5.5
- provide metadata describing their digitized objects. Six metadata elements are required when describing these objects:
    - title
    - creator
    - subject
    - date created
    - language (where applicable)
    - identifier
- establish a disaster-recovery plan as well as data backup procedures

Accessibility

CCO requires its fund recipients to create Web sites that can be viewed by all Canadians, regardless of physical disabilities.

- CCO funded recipients **are required to** create sites that either conform to accessibility guidelines as defined by the W3C's Web Accessibility Initiative (WAI) Guidelines 1.0 (http://www.w3c.org/TR/WCAG10/) or make available alternate versions that comply with these guidelines.

Several resources can be of help to site developers to meet these conformance requirements. For example, WAI provides a checklist: "Checklist for Web Content Accessibility Guidelines 1.0" at http://www.w3.org/TR/WCAG10/full-checklist.html

- It is **recommended** that fund recipients adhere to the practices described in the Accessibility Tip Sheet developed by the Canadian Heritage Information Network for the Canadian museum community. See: http://www.chin.gc.ca/English/Digital_Content/Tip_Sheets/Accessibility/index.html

File Naming and Directory Structures

A well-considered set of file-naming practices offers multiple benefits both in the immediate and the long term. A file-naming convention helps to manage digitization workflow while a well-planned directory structure will improve the management of digital assets. It will enable portability of objects, better discovery by search engines and increase visibility to users.

- CCO **requires** that fund-recipients develop and document file- and directory-naming rules and conventions for CCO-funded projects.
- It is **recommended** that file names and directory structures be descriptive and intuitive for both site developers and users.

Consistent file-naming and directory structure conventions can also provide users with 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. It can also help ensure that objects are not overwritten where they are used for aggregating purposes (for example, collections of thumbnail images).

- It is **recommended** that CCO-funded projects 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.

Types of Content and Format Standards

The vast majority of CCO-funded projects will result in content that will be viewed by users on browsers over the Web. Since the Web browser is the intermediary between users and Web sites, it is important that all content and all files in which the content resides are built on standards supported by the various browsers available to users.

As the Web has grown in complexity and evolved to provide developers with more powerful tools, an increasingly larger number of file formats has emerged. Some formats are non-proprietary and recognized as open standards in their categories, while other formats are proprietary and subject to commercial licensing or use agreements.

Proprietary formats require the use of proprietary software to create, distribute, manage and/or view or experience the content available in this type of format.

- To encourage the development of content accessible to the greatest number of people, CCO **requires** the use of non-proprietary formats by fund recipients in the preparation of projects, or the implementation of a strategy to allow users to experience the content at no charge in cases where proprietary formats are used, i.e. using freely-available plug-ins or media players

Types of content formats addressed in the guidelines are: 1) text 2) still images 3) moving images/video, 4) sound 5) animation and, 6) databases.

Text

Text-based content for the Web is stored either in a database or in files on a Web server, and delivered to a browser.

- CCO **requires** that text-based content developed for delivery to a browser be displayed as HyperText Markup Language (HTML) or Extensible HyperText Markup Language (XHTML). The latest version recommended by W3C (currently HTML 4.01) or the latest version of XHTML (currently 1.0) **must** be used. HTML and XML documents **must** be validated against a published DTD and/or using services such as those freely supplied by the W3C and other organizations. Sources from the W3C:
    - 1) HTML 4.01 - http://www.w3.org/TR/html401/
    - 2) XHTML 1.0 - http://www.w3.org/TR/xhtml1/

- Text-based content can also be developed in a proprietary format for viewing or printing with the use of freely-available plug-in software over the Internet. An example of this type of format is Adobe's PDF format, which requires the use of the Adobe PDF viewer. Content prepared in such formats **must** also be made available in an alternate format, such as HTML, that is accessible to individuals with disabilities.
- While all text-based content must be displayed as HTML or XHTML as described above, such content can also be delivered in a second format such as RTF, ASCII or delimited text files. Content produced in such formats is acceptable if the files are intended for download, storage or manipulation by users outside of the browser environment and if they are not meant to be a substitute to creating content in HTML or XHTML.
- To promote the long-term viability of text-based content, CCO **recommends** that fund recipients create and manage content in a structured format suitable for delivery as HTML files. In most cases storing text-based content as HTML, XHTML, XML or ASCII text will be the most appropriate option.

This will reduce the dependency on any content creation or management software and on the server environment on which it is stored. This, in turn, will extend the lifespan of the content.

- Text-based content can also be developed and managed in a database or content management system. CCO recommends that fund recipients choose a system that will allow the content to be exported in a standardized format such as a delimited text format. This will ensure that the content managed in such environments is not completely dependent on a single application or platform and, at a minimum, that it can be migrated to new environments with greater ease.

Note: The requirements listed above do not apply to text found within still images, moving images or animation files.

Still Images

The development of the Mosaic browser and server in 1994 made it possible to incorporate still images, or graphics, into Web pages. This feature has arguably contributed more than any other to the growth in the Web's early popularity. Today, graphics are ubiquitous on the Web but are nonetheless subject to standards, as they can exist in a variety of formats:

- CCO **requires** that fund recipients use images in GIF or PNG formats in the case of line-drawn graphics or in JPEG format in the case of photographs when displaying them on their Web sites.
- In the case of projects involving the digitization or creation of photographs or visual art, or the original creation of digital images leading to the creation of collections of images, images must be initially created and stored using the TIFF or PNG formats, which are uncompressed formats. These formats are suitable for the initial capture of the images, whether this is the result of a scanning process or of a digital creation process. More importantly, it is most suitable for preserving and archiving content for future use.
- TIFF or PNG formatted images **must** be stored in an environment such that fund recipients can access them at a future date. These files **must** therefore be stored on a server (subject to regular backups) or on CDs.
- Depending on the project, fund recipients **may** provide images in a second format beyond those mentioned above for the purpose of allowing downloading of images.
- CCO **may** grant exemptions to the above requirements in certain situations where the use of proprietary solutions is needed to meet project objectives. This could be the case, for example, where geographic maps are being digitized. In such cases, specialized solutions could be chosen to provide users with greater flexibility in viewing the digitized objects. Fund recipients must provide a justification for the use of such technologies in the context of their project objectives. Fund recipients must allow the content to be viewed or experienced at no charge, e.g. through the use of a freely-available plug-in or media player.

Moving Images / Video

As issues surrounding bandwidth, access to multimedia computers and tools to create digitized video have been resolved, more and more Web site producers are integrating video into their projects. Online video files require the use of specialized players in order for users to view them.

- Video and moving images files **must** be prepared in a format that can be played by a freely-available player such as Apple's Quick Time Player, Microsoft's Media Player, Macromedia Flash Player and Real's RealOne Player. Files can be saved in formats such as Quick Time, Real Media, RAM, AVI, Windows Media, SMIL and MPEG file formats.
- To promote accessibility of the content, fund recipients **are required to** provide alternate, text-based

summaries of video files that allow any user to understand the essence of individual clips.

Video files can be developed to be delivered as streamed video or as files to download prior to viewing.

- If video files are prepared for delivery in a high-bandwidth environment, an alternate low-bandwidth version **must** also be prepared and provided to users.
- Some streaming systems allow producers to prepare a single video file that can be played at various streaming speeds that correspond to multiple Internet access speeds. In cases where such systems are used (examples include Real's SureStream and MS Windows Intelligent Stream), a single version of a video file is sufficient.
- Where codecs are used to produce digital video files in order to compress the content, fund recipients **must** ensure that they are included in a platform (such as the Apple, Real or Microsoft platforms) or freely-available for installation by users. A link to codecs must be provided for users who need to download and install them.

Sound / Audio

As with video, sound and audio files have gained in popularity on the Internet in recent years. Sound files require the use of players for users to hear the content they include.

- Sound files **must** be prepared in a format that can be read by at least one of the freely-available players and can include RM, WAV, AU, MIDI, FLV and MP3 file formats.
- To promote accessibility of the content, fund recipients **are required to** provide alternate, text-based summaries of audio files that allow any user to understand the essence of individual clips.

Sound files can be developed to be delivered as streamed audio or as files to download prior to hearing.

- If sound and audio files are prepared for delivery in a high-bandwidth environment, an alternate low-bandwidth version **must** also be prepared and provided to users.
- Some streaming systems allow producers to prepare a single audio file than can be played at various streaming speeds that correspond to multiple Internet access speeds. In cases where such systems are used (examples include Real's SureStream and MS Windows Intelligent Stream), a single version of an audio file is sufficient.
- Where codecs are used to produce the audio files to compress the content, fund recipients **must** ensure that they are implicit to a platform (such as the Apple, Real or Microsoft platforms) or freely-available for installation by users. A link to codecs must be provided for users who need to download and install them.

Animation

Collaborative work in standards bodies, such as the W3C, as well as industry organizations regarding animation is such that the development of animation currently involves the use of both open standards as well as proprietary components.

- When creating animation, fund recipients are **required** to develop the animated content in a format that can be played with the use of plug-in software freely-available to users.
- A link to the software, for those users who need to download and install the software, **must** be provided.
- To promote accessibility of the content, fund recipients are **required** to provide alternate, text-based summaries of animation files that allow any user to understand the essence of individual animation files.

It is **recommended** that animation be created on a platform based as much as possible on open standards such as the W3C's Scalable Vector Graphics 1.1 specification, SMIL or MPEG 4 standard.

Databases

Many Web sites, particularly complex or large-scale sites, make use of databases to store content. Users therefore never interact directly with the database but rather do so through the browser, which will show the results of a database query or allow the user to accomplish certain functions via a form.

The database standards and requirements below are aimed at ensuring interoperability and preservation of data.

- Fund recipients **are required to** use a database that supports open standards such as SQL and XML. The database solution **must** allow fund recipients to import and export the data into a range of commonly-used formats that are independent of any particular platform rather than dependent on a proprietary one.
- Fund recipients **are required to** document their database designs and structures. Documentation about the design, the rationale behind it and decisions for normalizations, etc., will serve several purposes. These include facilitating possible migrations, data recovery and preservation; not to mention allowing others to better understand the database itself in cases of future maintenance and software updates.
- Fund recipients building databases to store metadata are **encouraged** to follow the Open Archives Initiative Protocol for Metadata Harvesting. This will allow access to data by harvesters such as those of cultural information gateways and metadata repositories.
  Source: http://www.openarchives.org/

Markup and Scripting Languages

Markup Languages

The universal language on which Web sites are built and understood by computers accessing the World Wide Web is HyperText Markup Language (HTML). As the W3C puts it, HTML is the "lingua franca" for publishing hypertext on the Web.

The fundamental purpose of HTML is to structure text and other objects integrated into Web pages such as graphics and video. While many developers use HTML to define presentational characteristics of their pages, style sheets have now emerged as the preferred method of describing the layout of a Web page. The following are requirements and recommendations regarding the use of HTML and style sheets.

- As mentioned in the "Text" section above, fund recipients **are required to** develop content for delivery to a browser in HTML or XHTML formats as recommended by the W3C.

- Fund recipients are **encouraged** to choose tools and technologies that produce HTML and Web pages that will be properly displayed on multiple browsers. CCO encourages fund recipients to incorporate the principles of Web site interoperability across browsers and platforms into their site designs.
- CCO strongly **encourages** fund recipients to develop Web pages that separate the overall structure

from presentation features, using HTML to determine structure of the pages and Cascading Style Sheets (CSS) to determine the visual presentation of content elements. Experience has shown that separating the structure of a document from its presentational aspects reduces the cost of serving a wide range of platforms, media, etc., and facilitates document revisions.

- Fund recipients should use Cascading Style Sheets, level 1 when developing content in HTML to determine the presentation and layout characteristics of their Web pages.
- Fund recipients should validate both HTML and CSS with the assistance of validators before publishing content to their public Web sites. The W3C provides access to an online CSS validator: http://jigsaw.w3.org/css-validator/; and to an online HTML validator: http://validator.w3.org/.

Client-side Scripting

Scripts are often incorporated into Web pages for a variety of purposes, including increasing the interactive quality of a page and extending the capabilities of the markup language. Client-side scripts depend on users' browsers to interpret them to accomplish their designated tasks. For example, a script could be created to alternate between two graphics when the user's cursor passes over a particular graphic with his/her mouse. This effect depends on the user's browser interpreting the script to switch between the two graphics when the mouse is "over" a specific graphic.

The most broadly used of the scripting languages are JavaScript, originally developed by Netscape, and JScript, developed by Microsoft. Shortly after the appearance of Internet scripting languages, the Switzerland-based standards organization, the European Computer Manufacturers Association (ECMA) proposed a standard for scripting languages known as ECMAScript. Both Netscape and Microsoft have developed new versions of their scripting languages that comply with this standard.

- CCO **requires** that fund recipients developing Web projects that include client-side scripts ensure that scripts are compliant with ECMAScript, version 3, also known as ECMA-262.
  The source for ECMAScript is:
  http://www.ecma-international.org/publications/standards/Ecma-262.htm.
- CCO **recommends** that fund recipients develop client-side scripts using either JavaScript version 1.5 or JScript version 5.5, which most closely comply with ECMAScript.
- Fund recipients **are required to** test pages on which client-side scripts are found using the last three versions of each major browser. This will ensure the largest number of users can view the pages correctly.
- Fund recipients should take into account the possibility that users will disable client-side scripts on their browsers. Key features of a site should function when scripting is disabled on a browser.

METADATA

CCO Metadata Requirements

One of CCO's objectives is to facilitate the discovery and increase the visibility of Canadian cultural content on the Internet. That is why, for projects funded by CCO, metadata is required. Metadata, or "data about data", is defined as structured descriptions of objects in real or online environments. Metadata, like a library catalogue, connects users to the information they want to find and use. In the online world, applying metadata can, among other things, help manage content, improve site navigation and manage rights associated with content.

We require the application of six metadata elements for all CCO-funded projects. Other metadata elements can be added to projects but that decision is at the discretion of individual organizations.

The application of a mandatory set of six metadata elements will help to ensure that sites funded by CCO can be found by Canadians and others interested in Canadian culture. Cultural gateway sites such as Culture.ca will be able to take advantage of the metadata to help users find exactly what they are looking for and thus, bring Canadian cultural content to broader audiences.

Annex A of this document provides an overview of metadata and what it can help you accomplish.

Common Metadata Element Set: Mandatory Elements

Fund recipients are **required** to apply the following six mandatory elements to describe their online resources:

Title
Creator
Subject
Date Created
Language (where applicable)
Identifier

Common Metadata Element Set: Recommended Elements

CCO also recommends (but does not require) that fund recipients apply an additional nine elements based on their relevance to each recipient's project. **These optional elements are**:

Description
Publisher
Contributor
Type
Format
Source
Relation
Coverage
Rights

Please read the accompanying document entitled CCO Metadata Application Guidelines for detailed information, guidance and best practices for each element listed above.

These nine recommended elements, together with the six mandatory elements noted above, comprise a core element set based on the internationally-recognized Dublin Core Metadata Initiative element set, adopted in 2003 as an International Standards Organization (ISO) standard.

CCO also recommends that fund recipients consider additional element sets if their project falls under one of the following four types of cultural resources:

1) learning resources;
2) audio-visual resources;
3) collective resources; and
4) resources supporting rights management.

The application of these sets by CCO fund recipients is optional. The list of elements for each of the four sets can be found in the CCO Metadata Application Guidelines. It also provides detailed information, guidance and best practices on each element.

Fund recipients may also apply additional metadata elements not listed in any of the elements sets provided by CCO. It is strongly recommended that additional elements be selected from existing international metadata standards. This will provide projects with greater long-term flexibility and a solid foundation for interoperability.

If you are using a metadata element set developed specifically for your organization, CCO requires that the six mandatory elements of the CCO comment element set be present in your metadata profile and that you demonstrate this adherence.

Application of the CCO Metadata Element Set

Identifying the Appropriate Objects to Tag

As important as identifying the purpose and actual elements to use is the question of what online content should be described within a Web site. Is the insertion of metadata tags on the main page of a site sufficient? Or on the first page of each main section? Should individual items, such as images, be tagged?

Since each Web site is unique, each will need to be evaluated individually to determine the correct level of depth. To determine this, CCO recommends that the following principles be applied to CCO-funded projects:

- The resources tagged should be resources for which there is sufficient context and meaning and are also worth listing in a search engine. A navigation button such as “Top of Page” would likely not be worth tagging but a digitized photo that is part of a collection would be worth tagging and thus made available as an individual object.
- If you are digitizing a collection, each item in it should be tagged individually.
- Apply metadata to resources in which users will likely be interested.
- Keep in mind that the more resources that are tagged, the greater the visibility your work will have.

Metadata and Format

Metadata can be stored and maintained within the resource itself (embedded metadata) or it can be stored and maintained independently of the resource (stand-alone metadata). For example, it is possible to store

some metadata about a digitized image within a header that becomes part of the image file during the digitization process. Alternatively, metadata can be stored in a database and linked to the digital resource.

- In cases where it is critical that metadata and content coexist, CCO **recommends** that the metadata be embedded.
- In cases where metadata is stored separately, the metadata **must** be accessible to data harvesting software. This will ensure that online resources can be indexed by other search or gateway sites with the use of metadata.

Expressing Metadata

- Each piece of metadata **must** be expressed in a standardized way to be read, searched, and exchanged by computer systems.

Metadata can be expressed in HTML (Hypertext Markup Language) or XML (eXtensible Markup Language). Simple metadata, such as unqualified Dublin Core, can easily be expressed in HTML. More complex metadata, such as qualified Dublin Core, can be expressed in HTML, but with limitations. XML has been used more successfully for highly-structured, complex metadata.

Rules for Individual Elements

Each element is subject to its own set of rules that indicate how its content and values should be expressed.

- CCO **recommends** that CCO-fund recipients review and apply the rules outlined for the six mandatory elements (and any other elements recipients opt to use as well) in the CCO Metadata Application Guidelines document. It provides detailed information, guidance and best practices for each element.

Where possible, CCO has applied ISO standards to ensure interoperability, that is, to ensure that CCO-funded sites express metadata values in the same way as other sites on the Web.

Staff Training

CCO-fund recipients should keep in mind that the quality of the metadata they produce will directly affect how successfully their content will be found. The higher the quality of the metadata, the more likely their online resources will be discovered by users. It is therefore important that individuals responsible for metadata creation be properly trained to carry out this task.

- CCO **recommends** that fund recipients ensure that individuals responsible for the creation of metadata be trained to carry out this task.

In addition, organizations should consider the use of subject specialists or professional indexers and cataloguers.

Steps in Developing Metadata

- Organizations should define their metadata requirements early in the project planning cycle.
- Metadata creation should be integrated into the digital production and development cycle rather than be left to the end of the process.

This will allow organizations to establish requirements and plan for appropriate resources and training at the beginning of project development.

Project Management Practices

Production Equipment and Facilities

Fund recipients **are encouraged** to use industry standards and appropriate equipment and techniques for the capture and storage of digital resources. It is important to consider whether a 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, skill-set available in house etc.

Preservation of Digital Resources

CCO recognizes the many difficulties inherent in the long-term preservation of digital information. CCO **encourages** funding recipients who are concerned about the long-term preservation of digital materials to contact CCO directly for assistance and / or to arrange for deposits of collections material with one of Canada's federal cultural institutions as stored copies, to allow for future authentication, backup and archival activities.

Backups and Disaster Recovery Planning

- CCO funding recipients **are required to** establish a disaster recovery plan as well as data backup procedures. These procedures should include a process to regularly backup data and store the backup medium (tapes, data CDs, etc.) in an area other than the site-hosting location.

Appendix A
Metadata – A Primer

What is Metadata?

Metadata is information about an object or a resource that is organized in a structured fashion. The best analogy for understanding the nature of metadata is a typical library catalogue. In a catalogue, every book is described by a record that includes the author, title, subject, date of publication, etc. All catalogue records are structured in the same way and with the same fields of information. Records give users enough information to figure out which resources, books and journals (from among the hundreds of thousands in the library) will meet their needs, and where the resources are located in the library building. Moreover, the catalogue provides several paths to information by allowing users to search for resources with similar characteristics, such as all books by one author or on one subject.

In the digital realm, metadata, like a library catalogue, provides some basic information about digital resources. Users can search metadata records to find digital resources much like they would search a library catalogue to find a particular book. Without structured metadata, users are left sifting through mounds of irrelevant information to find what they are looking for.

The application of metadata is the strategy chosen by CCO and many digital content experts to produce structured information that makes it easier for users to locate the digital resources they are looking for.

An Example of Metadata

If, for example, a high school student was preparing a presentation on the history of women in hockey in Canada. She would like to include photos of early women's hockey teams as part of a presentation to her class. Her teacher has suggested she look at the Library and Archives Canada's Web site as a possible source of photos.

The site includes a section called Images Canada (http://www.imagescanada.ca/), an online search tool that provides access to thousands of digitized photos from several Canadian cultural institutions. Using the Images Canada search function, the student types in "women, hockey" as her keywords. The search returns nine records, each of which includes a link to a digitized version of a photo.

Each photo is accompanied by detailed information, which allows the student to properly describe the photos she chooses for her presentation. One of the photos is accompanied by the following details:

Title: Women's hockey team, University of Alberta, Edmonton, Alberta.

Image: View full size image

Source: Glenbow Archives

Subject: Edmonton, Alberta - Sports | Edmonton, Alberta - Personalities | Edmonton, Alberta - Newspapers | Hockey

Description: Back row L-R: H. Moore; M. Russel; R. Wood; J. Hennesey; H. Barclay; W. Matthews, coach. Front row L-R: D. Whiteman, manager; L. Macgregor; T. McQueen, captain. From "Evergreen and Gold" page 83.

Date: 1921

Type: Image

Format: Jpeg

Language: English

Unique ID: NA-4030-3

Coverage: Canada - Alberta

Rights: Copyright

This information represents metadata. Each of the nine records includes similar information. Metadata allows the student to find the photo because both “women” and “hockey” are found in the title of the photo. It also provides details such as the year the photo was taken, where the hockey team came from and a description of who is in the photo. It also includes important information on the use of the photo through its description of copyright. This information assures the student that she has permission to use the photo for her school project.

In this example, metadata serves several purposes: 1) it allows the student to find photos for her school project; 2) it provides detailed information about each photo that she can use in her presentation; and 3) it provides the cultural institution, namely the Glenbow Archives, with the means of controlling how the photo is used as described in its copyright policy.

Metadata Structure

As mentioned earlier, metadata expresses information about resources, structured into categories such as title, type and source. These categories of information about an object are called metadata **elements**. In our example above, twelve elements are listed. A grouping of elements identified as an appropriate assortment of characteristics is called an **element set**. In the case of the Images Canada example, the element set consists of: 1) title; 2) image; 3) source; 4) subject; 5) description; 6) date; 7) type; 8) format; 9) language; 10) unique ID; 11) coverage; and 12) rights.

There are advantages to organizations agreeing to use the same element set. In the case of Images Canada, the catalogue of photos is actually populated by the individual collections of several institutions. The Images Canada collection exists in large part because the partnering institutions agreed to use the same element set to describe their photos. The seamless sharing of information in this way is what is meant by interoperability and is possible because of the wide acceptance of certain element sets.

The Images Canada metadata element set is based on an international standard: the Dublin Core Metadata Initiative (DCMI) element set. The Initiative is a voluntary group that brings together individuals and organizations interested in creating a basic set of metadata elements applicable in any content on the Internet. The complete DCMI standard includes 15 elements.

Another element set that has emerged in recent years is the one created by the Institute of Electrical and Electronic Engineers (IEEE) for use in the context of e-learning resources. The Learning Object Metadata Standard that this organization's members have developed is composed of 76 elements.

Several other element sets have been developed by international standards organizations to describe various types of digital resources. Each set includes specialized elements that describe particular characteristics of digital resources.

Beyond element sets, other issues include the format in which an element's value is displayed (e.g., do we write a date as "January 1, 2004", "01-01-04" or "1/1/04", etc.) as well as the syntax used to write metadata (is it included in the HTML code of a Web page or is it stored in a database or even in XML files?). Standards in each area provide direction to address such problems.

Uses of Metadata

Metadata has no use on its own and is generally meant to be read and interpreted by a computer. The results of the computer interaction are then made available to the user and take a form that is more easily understood. In other words, it requires a software application, such as a content management system - a metadata-enabled search engine or catalogue software - to make it usable and useful.

There are many examples of organizations using metadata to enhance the functionality and value of their Web sites on the Internet. There are also benefits to metadata that go beyond the individual site, such as providing value to sites that aggregate information. These are the search engines, directories, gateways and specialized portals that connect users to the information they seek.

Searching for Content

Simply making content available on the Internet does not guarantee that users will find the material in which they are interested. Metadata can provide different ways for users to look for information on a Web site.

There is momentum building in favour of metadata as one of the better tools to help sort through the mass of information published everyday on the Web. In the competitive and fast-paced world of search engines, techniques that harvest metadata and increase the efficiency of searches will undoubtedly surface and provide greater advantage to some of the existing engines.

In the immediate, metadata is perhaps of greatest use in directories, gateways and specialized portals. Sites such as Culture.ca use metadata extensively to assist users in locating Canadian cultural resources. More widespread application of metadata will only increase the gateway's efficiency in giving content greater visibility.

An example of a specialized portal using metadata is the Movie Review Query Engine (http://www.mrqe.com), which provides links to reviews of films and includes data such as the publication in which the review appears, the language of the review, the film's rating, location on the Web, etc.

Developing metadata today ensures you will be better positioned in tomorrow's Internet.

Collaboration and sharing

Canada's online cultural resources are held in repositories, databases and web servers across the country. Well-structured metadata records make it possible for users to search across multiple collections using online gateways and repositories, establish relationships, and create virtual collections from materials that are distributed across several repositories. The Images Canada (http://www.imagescanada.ca/index-e.html) collection of photos we used in our earlier example clearly demonstrates the advantages of collaboration.

OTHER USES BY ORGANIZATIONS

Metadata can be applied to great effect within a Web site to meet the needs of an organization. It can, for instance, help support a broad range of functions used by visitors to a site.

Navigation: Metadata can be used to create navigation menus and help users to navigate a site or a collection.

Support for life cycle of content and copyright: Metadata can be used to document the layers of rights and reproduction information that exist for original resources and their multiple versions.

Other uses within an organization can include: 1) the management of content versions of documents (for example, where low and high-resolution versions of the same works of art are created); 2) data stability (documenting technology requirements for some resources); and 3) protecting investments (metadata can add to the value of resources by increasing their usefulness).