———— **A R T I C L E S** ————

Collection Level Description - the Museum Perspective

Heather Dunn
Professional Programs
Canadian Heritage Information Network

Introduction

Many organizations -- libraries, archives, museums, government agencies, and others -- have content that they need to make accessible on the Internet. Museums have long had computerized collections management databases, and many of them are offering item-level meta-data to the public on the Web. But the content of museums� collections databases is invisible to search engines -- so what is the means for resource discovery on the Web?

The creation of Web pages containing collection-level description (meta-data describing the collection) is a solution that would facilitate resource discovery. However, there are problems associated with the creation of these descriptions, whether they are manually written, or automatically generated from the item-level data. What is collection-level description in the museum context? How can these descriptions be created? What terminology should be used? We need to take a closer look at these issues if we wish to use collection-level description to assist in allowing access between organizations, across disciplines, among resources with different types of content, and for audiences with varying levels of expertise.

1. Defining "Collection-Level" in the Museum View

What is collection-level description in the museum context? Even within a single museum, the concept of a "collection" may have many different interpretations. Consider the following definition of "collection" from the ArtLex Lexicon of Visual Art Terminology:

> "collection -- An accumulation of objects�. Collections can be formed around any of a variety of parameters. They may be centered upon a medium or technique, a certain period or group of artists, or a subject, for instance; or they may be encyclopedic, as can be the entire collection of a large museum. Museums typically have both permanent collections and traveling collections. Also see accession, deaccession, donation, gallery, and patron". [1]

In the museum view, a collection may consist of the entire holdings of a particular museum. Or, a collection may be a discrete part of a museum�s collection; it may be centred on some type of similarity between the items in the collection. For example, a museum collection may consist of the works of a particular artist (e.g., the Monet collection) or group of artists, a particular medium or technique (the print collection), or a certain discipline (the ethnology collection). Museums also define collections in administrative terms: for example, the collection of a particular donor (the Barnes collection), or suited for a particular purpose (the education collection).

The concept of "collection" is different again in the context of a collaborative digital resource built from the content of many museums. In a collaborative digital resource, the concept of "collection" becomes very fluid; it can extend beyond the physical walls of individual museums and allow users to combine and re-combine objects at will. For example, a user of on-line resources might wish to bring together all the works of art created by one artist, regardless of where they are physically housed. A "virtual exhibit" can also be considered a collection, bringing together individual objects housed in many physical locations.

Researchers wishing to access information about these collections on-line may be interested in any of the above interpretations of "collection". Museums cannot predict what their users will consider to be a collection, in what language or terminology they will request the data, or what level of information they need (is it for an elementary school project, or a Ph.D thesis?) Ideally, collection-level description for resource discovery should provide access to many interpretations of "collections" that are dynamically created by the user.

2. Why Is Collection-Level Description Important?

Many museums wish to provide public access to their collections databases over the Internet. One of the main reasons for creating collection-level descriptions is resource discovery of object-level information held within databases. Database contents (information and images about individual museum objects, for example) are invisible to Web searches, and it is often impractical to create a Web page for each item in the database. Unless the contents of databases are described on the Internet at the collection level, users will not be able to find the data in a Web search. Individual museums can use collection-level description to help Web users discover the item-level information held in their database resources.

Some museums have created Web pages for individual items in their collection. For example, the Web site of the Metropolitan Museum of Art, at <http://www.metmuseum.org/collections/index.asp > provides access to a large collection. Collection-level description could be used in such cases to facilitate resource discovery of museum Web pages containing item-level descriptions or images. It can be used to help a searcher find a general class of items, even though the museum Web site contains only references to specific instances of that class. For example, an art museum may have a Web site containing text and images about their collection of works by Monet. Although a Web search for "Monet" may find the museum�s Web pages, a search for "Impressionist" or "French Artists" may not. If the museum�s Web site contains only images, it will not be found in a Web search at all. Appropriate terms can be added to the collection-level description to ensure that the museum Web site is found, whether the researcher

uses broad or specific terminology in their search.

It is also possible to use collection-level descriptions for resource discovery within distributed resources. Collection-level descriptions of resources created by individual museums (for example, descriptions of a museum collection, virtual exhibit, professional resource, etc.) can be contributed to a centralized location (e.g., a subject gateway or search engine) to be searched by users. When the users find a description of a resource that meets their needs, they can link to it and explore it in detail.

More and more importance is being placed on data sharing among organizations -- museums, archives, and libraries, and government agencies, as well as the corporate world, are striving to be "interoperable" at the local, national, and even global level. Collection-level description is also important here. For example, a museum might have a collection of Impressionist paintings, while a library holds a collection of books on the Impressionist movement and individual Impressionist artists. A museum might hold objects relating to the early settlement of a community, while governmental agencies or archives might hold statistical studies that are related (e.g., a census for the period). A researcher would not be able to see the relationship between the collections of the museum, archives, library, and government agency unless there were collection-level descriptions for each.

So, collection-level descriptions facilitate cross-disciplinary, multi-level access to Web and database resources for a diverse audience. But how would these collection-level descriptions ideally be created to fulfill these objectives?

3. What Are the Ideals for Collection-Level Description? Are They Attainable?

Ideally, collection-level descriptions would be created following a well-designed standard which had been adopted globally, across disciplines and was suitable for resource description at the object or collection level. They would be automatically, dynamically created according to user requirements. They would be multi-lingual, and provide semantic links between object and class, between professional and public terminology. This is obviously not the reality, but how far are we from this ideal? What steps are being taken to make it happen?

The Consortium for the Interchange of Museum Information (CIMI) [2] has made some important advancements in the field of standards for museum resource description. Phase 1 of the CIMI Dublin Core Testbed Project [3] was undertaken in 1998 with the goal of testing "assumptions related to the flexibility and simplicity of the Dublin Core element set, and its suitability and readiness for deployment". Seventeen CIMI member organizations worked to create object-level descriptions using the Dublin Core standard, and identified issues surrounding the functionality of Dublin Core for resource discovery on the Internet. One of the problematic topics that was raised as a result of this project was "characterizing resources as either item-level or collection-level -- i.e., determining the unit of analysis for description such as with an exhibition, a collage of photographs, or other aggregated objects" (Section 5.2 Issues). Phase 2 of the Dublin Core Testbed Project [4], which began in 1999, includes the publication of a "Guide to Best Practice" for museums using Dublin Core, and an "examination of Resource Description Framework (RDF) as an effective method for enabling interoperability between applications that exchange meta-data.

RDF, an emerging standard of the World Wide Web Consortium (W3C)[5], is "a foundation for processing meta-data; it provides interoperability between applications that exchange machine-understandable information on the Web" [6]. Among other applications, RDF can be used in resource discovery, cataloguing, and collection-level description. The W3C has developed a model for representing RDF meta-data, and has recommended the use of Extensible Markup Language (XML) as a syntax for encoding this meta-data.

The development of these standards for the creation, processing and encoding of meta-data is a vital step toward the goal of achieving cross-domain interoperability. Although they have not yet been widely utilized by museums, and there are still many issues that need to be resolved, their validity for museum data has been demonstrated through the CIMI Testbed project.

Despite the great advancement in standards for resource description over the past few years, the terminological issues seem more daunting than ever. Ideally, the collection-level description should provide access to both general and specific requests, regardless of the knowledge level, discipline, data requirements, and language of the user. Knowledge representation tools such as thesauri can accomplish some of these goals. In order to bridge the semantic gaps between the language used in the collection-level description and the language used by Internet users (who come with varying levels of subject knowledge), thesauri can be employed with the search engine. Taking advantage of the associative, hierarchical, and equivalency relationships of a thesaurus will allow easier access in that the terminology used by the searcher does not have to match that of the resource description.

Part of the appeal of using RDF as the foundation for meta-data processing is that it allows use of the XML namespace facility[7] The namespace facility allows the documentation of restrictions and definitions for an organization�s meta-data. For example, a museum that is using the Art & Architecture Thesaurus [8] as the controlled vocabulary for object names can declare this in their namespace. This machine-readable namespace declaration will eventually ensure that the meta-data is "understood" and processed as intended by its creators. In theory, by reading the namespace of an organization, a resource discovery tool will be able to determine the precise meaning of each of the organization�s meta-data elements, have access to each controlled vocabulary used in the organization�s meta-data, etc.

4. The Reality: Collection-Level Description in CHIN Resources

With these ideals in mind, museums are progressing toward the goal of using collection-level description to achieve interoperability. However, there are many problematic issues that need to be resolved before this can be accomplished. What terminology should be used in collection-level description to ensure access by both the generalist and the specialist? How can we provide access in multiple languages? How can we create linkages between the terminology used in the collection-level descriptions and the object-level descriptions? How can we accomplish interoperability when the emerging standards are still moving targets? Some of the initiatives of the Canadian Heritage Information Network (CHIN) [9] can be used as an illustration of the use (and potential problems) of collection-level description for resource discovery, and of temporary solutions that can be employed until the meta-data and terminology standards catch up with user requirements.

Some of the most problematic issues have to do with the terminology used in resource description. Museums often use highly specialized terminology to describe their collections, whereas Internet users with no subject experience may use very general terms. Conversely, museums may create collection-level descriptions using very general terms, to the frustration of the user who is searching for a very specific item. It is important that the terminology used in the collection-level description is specific enough to allow users to decide whether they have found an appropriate

resource, but general and descriptive enough so that people from a wide range of disciplines and knowledge levels can discover the resources. This is easier said than done, as the discussion below will illustrate.

CHIN has had considerable involvement with resource description in the museum context; for over 25 years, museums across Canada have been contributing object-level meta-data (sub-sets of their collections management information) to a collective resource managed by CHIN. This growing, collective resource, now called Artefacts Canada, contains data on over 2 million objects housed in Canadian museums, and is accessible to the public on the CHIN Web site. CHIN has created links between the object-level meta-data in Artefacts Canada and the collection-level descriptions found in another CHIN product, *The Great Canadian Guide*.

Museums contributing object-level data to *Artefacts Canada* do not use a common terminological standard; many museums do not use a standard at all. Indeed, no vocabulary standard exists that would meet the needs of all museums. A wide range of vocabulary is used by the museums, from very specialized to very general terms, in both English and French. Some museums use a classification system, such as the *Revised Nomenclature for Museum Cataloging* [10], in addition to object names; some do not use any classification system. To mitigate this problem, the Getty�s *Art & Architecture Thesaurus* <http://shiva.pub.getty.edu/aat_browser/> (with CHIN�s addition of the 2600 most commonly-used French terms) has been integrated with the *Artefacts Canada* search engine. This enables the user to enter a search term such as "painting", for example, and obtain results which include objects that are catalogued as "watercolour" (a narrower term to painting), and "peinture" (the French term for painting). Although the *Artefacts Canada* search engine works quite well in itself, the individual *Artefacts Canada* records are invisible to Web searches. One solution to this problem is to provide Web pages with collection-level descriptions for each of the museums, and to link from the collection-level description to individual instances of that museum�s collection. CHIN has accomplished this through the collection-level descriptions in *The Great Canadian Guide*.

The Great Canadian Guide is another CHIN resource that has been produced in collaboration with the Canadian museum community. The *Guide* is an on-line gateway to over 2400 Canadian cultural institutions and attractions; museums use an on-line form to provide and update basic information on their exhibits, hours, location, etc. Each museum provides a short collection-level (scope of collections) description using free text, in either English or French (or both). They also select terms that represent their collection (e.g., Clocks or Time-keeping Devices) from a controlled vocabulary list.

Because the museums are required to use a controlled vocabulary in describing their collections, the terms they use are standardized between institutions. Another reason for using controlled vocabularies for collection-level description in the *Guide* is that the terms can be automatically mapped to language equivalents so that the *Guide* information can be searched and displayed in either French or English.

The free-text portion of the collection-level description also serves an important purpose. In describing a collection, the significance of certain items in the collection may be higher than others, and should be highlighted for the user. For example, a museum may use very general terms to describe a collection of Impressionist art, but it is likely important to the user that the collection contains a work by Monet. The free-text portion of the description allows the museum to highlight individual items that cannot be described using the broad controlled vocabulary.

One of the features of the *Guide* is that it allows users to link from the collection-level description of a museum�s collection to the corresponding object records in *Artefacts Canada*. For example, if a user does a search on the Internet for "furniture", he may find the *Guide* page for the Provincial Museum of Alberta, because the Museum has used the term "Furniture or Furnishings" as one of its collection-level descriptors. The user can click on the term, and be presented with all the Museum�s chairs, tables, etc. as they are found in *Artefacts Canada*.

In this case, the collection-level description has been an effective means of resource discovery -- the user was able to find the resource through the collection-level description, and then investigate the individual objects within the collection. However, the linkages between the *Guide*�s collection-level description and the individual instances in *Artefacts Canada* are not always automatic. Museums seldom use standardized classification terms such as "Furniture or Furnishings" in their object records, and the *Art & Architecture Thesaurus* does not always make the link between the classification-level term and all the possible members of that class. Therefore, CHIN has had to create associations between the broad classes describing collections in the Guide to individual members of that class in *Artefacts Canada*. For example, when a user clicks on a collection-level descriptor in the *Guide* (e.g., "Clocks and Time-keeping Devices"), the search that is launched in *Artefacts Canada* has been manually supplemented with terms such as "watch", as this relationship is not included in the *Art & Architecture Thesaurus*. Although the terms used in the *Guide*�s collection-level descriptions are from a controlled list, and the *Artefacts Canada* search engine is assisted by the associative, hierarchical, and equivalency relationships of the *Art & Architecture Thesaurus*, there is still a semantic gap between the two levels of description.

Another problem with this model is that an Internet searcher will not be able to use specific terminology in a Web search. For example, if the user searched for "watch" on the Internet, the Whyte Museum entry in the *Guide* would not appear, as the specific term, "watch" is not in the collection-level description; the user would have had to know to search for "Clocks and Time-keeping Devices". Another problem is that the linkages between the resources are not dynamic. For example, if the museum adds records to *Artefacts Canada* that have not been included in the list of expanded search terms defined by CHIN, the linkages from the class to the member will not be there. If a museum removes all of its "Clocks and Time-keeping Devices" from *Artefacts Canada* without changing its collection-level description in the *Guide*, there will be a dead link in the *Guide*.

In attempting to link between object-level descriptions and collection-level descriptions, CHIN has chosen to have museums write collection-level descriptions for their collections, and is trying to develop methods of bridging the gaps between the search terms used by Internet users and the terms used by museums in their collection-level descriptions. As we have seen, this is difficult because the knowledge tools must be able to link from the class to the object (and vice versa), and also create the connections between specialist and generalist terminology in multiple languages. Museums are using a wide variety of controlled vocabularies such as classification tools and thesauri (and some are not using any); we need to find a way to bring these tools together. Multilingual knowledge tools are the goal, but at the very least, there is a need for language equivalents at the classification level. No tool exists at present that will accomplish all this.

It may also be possible to work backward from the object-level data to dynamically create collection-level description. For example, if a museum has catalogued its collection using specific terminology, we may be able to run these specific terms through a knowledge tool that would determine the general class to which those objects belong. This seems harder to accomplish, but could be done dynamically to reflect changes in the content of the resource. More study needs to be done to determine if this is feasible.

CHIN has just begun to use collection-level description to enable access to distributed resources. *Learning with Museums* is a resource that makes

on-line educational museum content more accessible on the Internet. It has a distributed architecture, with the resources existing on the museums� Web sites, and only a meta-data record contributed to the central repository at CHIN. To participate in the *Learning with Museums* project, museums create meta-data for each of their on-line resources (virtual exhibits, educational games, etc.), by using an on-line cataloguing tool created by CHIN and provided to members on the CHIN Web site. CHIN sends the meta-data back to the museum, and they embed it within their resource. CHIN can then harvest this meta-data periodically in order to update its meta-data repository. Users (teachers, students, etc.) can search or browse the *Learning with Museums* meta-data, and when they find a resource they are interested in, they can link to it, wherever on the Web it might be physically stored. *Learning with Museums* provides a description of the resource -- in some ways, this can be considered a collection-level description, as it is a description of the "collection" of items that has been brought together to form the virtual exhibition. Terminology used in these descriptions is based on a thesaurus of subject areas based on Canadian school curricula (e.g., Broader Term: Sciences; Narrower Term: Chemistry).

CHIN�s newest initiative, the Virtual Museum of Canada, is currently under construction. The Virtual Museum of Canada will use existing resources such as *Artefacts Canada*, the *Guide*, and *Learning with Museums* as building blocks, and will also enable Canadian museums to create new rich content to be included as distributed resources in the Virtual Museum of Canada.

Current thinking is that museums will use a cataloguing tool (similar to that used in *Learning with Museums*) to catalogue the content they are linking to the Virtual Museum of Canada. The cataloguing tool would allow them to provide CHIN with collection-level meta-data on the virtual exhibit, photo gallery, etc. that they have created. In addition to the collection-level descriptions that museums will provide to the Virtual Museum of Canada, museums will be required to submit object-level meta-data about the individual objects featured in their virtual exhibits to *Artefacts Canada*.

Again, there is a potential problem in linking the collection-level meta-data to the object-level meta-data. The Virtual Museum of Canada will include diverse types of content: collection-level meta-data pointing to virtual exhibits, Web pages with collection-level meta-data describing museums� scope of collections, and database records describing/illustrating individual museum objects. The challenge is to find a way to enable users to find individual items in a collection or virtual exhibit, as well as the entire exhibit/collection.

We are currently exploring the options for solving this problem. Library classification systems built into the search engines might offer part of the solution, if they are able to fill the semantic gap between our collection- and object-level descriptions. As well, we might be able to build linkages into the meta-data. For example, we could require that when a museum submits meta-data about a new virtual exhibit that is to be added to the Virtual Museum of Canada, the museum might also be required to identify each object featured in the exhibit with a unique identifier that already exists in the object�s record in *Artefacts Canada*.

The Virtual Museum of Canada will eventually allow cross-domain searching to the holdings of the National Library of Canada and the National Archives of Canada. Collection-level descriptions and RDF will doubtless play a large part in achieving seamless interoperability between these three disciplines.

Conclusion

Although there are still many challenges to be faced in using collection-level description to facilitate museum resource discovery on the Internet, advancements are being made through the development of standards such as Dublin Core and RDF, and the increasing use of knowledge tools such as the *Art & Architecture Thesaurus*. It is important that the museum, library, and archival communities work together to ensure that these developments lead to true interoperability and resource sharing on a global level.

References

[1] "Collection." *ArtLex Vistual Arts Dictionary*. July 26, 2000 <http://www.artlex.com/>

[2] Consortium for the Computer Interchange of Museum Information (CIMI). July 26, 2000 <http://www.cimi.org/>

[3] *CIMI Dublin Core Metadata Project Phase 1 Final Report*. Consortium for the Computer Interchange of Museum Information. July 26, 2000 <http://www.cimi.org/documents/meta_phase1_final_report.html>

[4] *CIMI Dublin Core Metadata Testbed Phase II*. Consortium for the Computer Interchange of Museum Information. July 26, 2000 <http://www.cimi.org/documents/meta_011899_pdII_final.html>

[5] World Wide Web Consortium (W3C). July 26, 2000 <http://www.w3.org/>

[6] *Resource Description Framework (RED) Model and Syntax Specification*. February 22, 1999. World Wide Web Consortium. July 26, 2000 <http://www.w3.org/TR/REC-rdf-syntax/>

[7] *Namespaces in XML*. January 14, 1999. World Wide Web Consortium. July 26, 2000 <http://www.w3.org/TR/REC-xml-names/>

[8] The Art & Architecture Thesaurus Browser. Version 3.0. J. Paul Getty Trust. July 26, 2000 <http://shiva.pub.getty.edu/aat_browser/>

[9] Canadian Heritage Information Network. July 26, 2000 <http://www.chin.gc.ca/>

[10] Blackaby, James R; Greeno, Patricia, and the Nomenclature Committee. 1988. *The Revised Nomenclature for Museum Cataloging: A Revised and Expanded Edition of Robert G. Chenhall's System for Classifying Man-Made Objects*. American Association for State and Local History, Nashville, TN.