Descriptive Metadata Guidelines

for RLG Cultural Materials

Contents

ACKNOWLEDGMENTS

Many thanks also to these individuals who reviewed the final draft of the document:

Caroline Arms, Rebecca Guenther, Mary Lacy, and Helena Zinkham, Library of Congress

Murtha Baca, Tom Burton-West, Erin Coburn, Jocelyn Gibbs, Billie Hackney, Patricia Harpring, Sally Hubbard, Deborah Lenert, and Joe Shubitowski, The Getty

Penny Baker, Laurie Glover, Valerie Krall, and Angela Sidman, Sterling and Francine Clark Art Institute

Matthew Beacom, Yale University

Alan Danskin, British Library

Stephen Hearn, University of Minnesota

Martin Kalfatovic and Suzanne Pilsk, Smithsonian Institution

Kris Kiesling, University of Texas at Austin

Marty Kurth, Cornell University

Rob Lancefield, Davison Art Center, Wesleyan University

Bill Landis, California Digital Library

Elisa Lanzi, Smith College

Dennis Meissner, Minnesota Historical Society

Clayton Redding, Princeton University

Sarah Stark, Bill Stockting and Sonia Ranade, National Archives (UK)

INTRODUCTION

RLG Cultural Materials (culturalmaterials.rlg.org) integrates collections from museums, libraries, and archives worldwide into a cross-community database of digital image, audio, and video surrogates and their associated descriptive metadata. This content is licensed to institutions for the support of teaching, learning, and research—all digital surrogates are copyright-cleared for education. Watermarked, low-resolution versions of "unrestricted" images generate revenue for contributors through RLG's Trove.net™ service (www.trove.net).

To ensure that the digital collections submitted to RLG Cultural Materials can be discovered and understood, RLG has compiled these Descriptive Metadata Guidelines for contributors. While these guidelines reflect the needs of one particular service, they also represent a case study in information sharing across community and national boundaries.

RLG Cultural Materials engages a wide range of contributors with different local practices and institutional priorities. Since it is impossible to find—and impractical to impose—one universally applicable standard as a submission format, RLG encourages contributors to follow the suite of standards applicable to their particular community. These guidelines therefore do not set a new standard for metadata submission, but rather support a baseline that can be met by any number of strategies, enabling participating institutions to leverage their local descriptions.

These guidelines also highlight the types of metadata that enhance functionality for RLG Cultural Materials. After a contributor submits a collection, RLG maps that description into the RLG Cultural Materials database using the RLG Cultural Materials data model. This ensures that metadata from the various participant communities is integrated for efficient searching and retrieval.

For contributors who are in the process of formulating policies for describing their collections—or simply trying to understand the choices and terminology—these guidelines provide an overview of key concepts and strategies. For contributors who already have a strategy in place for describing their collections, these guidelines will allow comparisons between local choices and RLG preferences. *Please remember that we are always happy to consider collections that do not meet all the outlined criteria.*

To contribute collections to RLG Cultural Materials, RLG requires participants to submit:

- A collection record describing the contributed content as a whole.
- Individual records describing the objects in the collection. Each record must have a corresponding digital surrogate, be it an image, an audio, or a video file.

Other resources from RLG

The Cultural Materials Alliance Content Manager (culturalmaterials.rlg.org/collectron/) on RLG's Web site is an essential tool for contributors to RLG Cultural Materials. Registering your collection there is the first step in alerting RLG to your intention to submit the material and gives you a way to provide metadata on the collection level.

These pages on RLG's Web site may also be helpful in preparing your RLG Cultural Materials submission:

- Contributing to RLG Cultural Materials
  www.rlg.org/en/page.php?Page_ID=523
  A high-level, process-oriented description of selection, digitization, metadata creation, and mechanisms of submission to RLG Cultural Materials.

- Preparing Digital Surrogates for RLG Cultural Materials
  www.rlg.org/en/page.php?Page_ID=224
  Information about how to digitize materials (still and moving images, audio), including filenaming conventions.

CHAPTER 1: TERMINOLOGY

Inconsistent terminology makes any conversation a challenge. These guidelines use a consistent terminology as defined here for the different aspects of descriptive metadata and the standards that govern them. To illustrate the different terms and their relationship, we can use a metaphor that likens a record to a case of bottles being labeled, filled, stored, and shipped.

Data fields & structure

- "Data fields" are the named units of information, often also referred to as "elements" or "categories." Data fields answer the questions, What discrete units of information do you want to capture? and, What are they called?

- "Data structure" is the organization or grouping of the data fields. Data structures answer the question, How are the data fields arranged into a record?

That is, data fields are the units of information a standard makes available for data capture. In terms of our metaphor, data fields are like bottles waiting to be filled. Data structures additionally articulate the order and hierarchical organization of the fields. Data field and structure standards may overlap with data format standards (see page 4)—the standard describing the units of information and their structure often also describes the preferred encoding of the fields into a file format. However, it is still advantageous to separate data fields and structure from data format. A standard's specifications for data fields—e.g., the Encoded Archival Description (EAD) tag library—may be implemented in a database for capturing and manipulating metadata, while the standard's data format specification may guide the way the database exports the data for exchange (e.g., the EAD XML DTD). Furthermore, data field and structure standards sometimes specify or suggest vocabularies for selected fields, overlapping with data content and value standards (see below).

Chapter 2 provides guidelines for data fields and structure, and Appendix A offers a list of relevant standards.

Data content & values

- "Data content" and "data values" are the information proper stored in the data fields. They answer the question, What goes into the data fields?

- Data content standards are the rules that guide you in filling a particular data field. They answer the question, What is the recipe for creating a term or a prose description?

- Data value standards are the thesauri or name authorities providing preestablished terms to populate a data field. They answer the question, What is the established term for a data field?

Chapter 1: Terminology

In terms of our metaphor, data content and data value standards establish what goes into a bottle. Data content standards specify the rules that mold data (such as a date) into predictable expressions (e.g., YYYY-MM-DD). While data content standards provide instructions for how to create descriptive metadata, data value standards provide preestablished terms. In thesauri, data value standards differentiate preferred terms (e.g., "asa-no-ha-toji") from other synonyms (e.g., "hemp leaf bindings") and provide context through a hierarchical organization of concepts; for an example see the Art & Architecture Thesaurus® (AAT). Name authorities provide the preferred term, as well as alternate and related terms, although typically without embedding them into a hierarchy.

Data content and value standards as interpreted and enforced by local cataloging guidelines greatly improve the consistency of the content in your data fields. This ensures that a query will produce a complete and relevant result—that all the records that should be retrieved are retrieved and records that should not be retrieved are not. This is true for an individual database but is especially important for collaborative projects such as RLG Cultural Materials.

Chapter 3 provides guidelines for data content and values, and Appendix B offers a list of relevant standards.

Data format

- "Data format" refers to the particular encoding of information contained within a file. It answers the question, How do the data fields and their content get saved as a digital file?

A data format is the particular encoding of information contained within a file. In terms of our metaphor, a data format corresponds to the crate in which the bottles can be stored and shipped. Different computing environments generate and store data in different data formats, which are often designed for a particular application. Data used by relational databases, for example, is stored in a different format than data used by spreadsheets. Data inevitably changes formats during its life cycle as it is transferred and transformed for other uses. System designers are well advised to take the fluidity of data formats into consideration during the initial development of new applications. Standards facilitate the maintenance and preservation of information within local systems, and they become even more important when data is shared with other applications. Transport systems require data to be in a format intelligible to both the sending and the receiving application. The most common file format for exchanging data in the cultural heritage community is XML (Extensible Markup Language).

Chapter 4 provides guidelines for data formats, and Appendix C offers two complete markup examples in XML.

Summary

Here is an example of how data fields and structure, data content and values, and data format specifications work hand in hand to establish a standards-based, shareable record:

- A data field and structure specification such as the Metadata Object Description Schema (MODS) Elements defines a discrete unit of information as "dateCreated".

- A data content standard such as ISO 8601 specifies the rules for formatting the date April 17, 1970 as "1970-04-17".

- A data format such as the XML expression of MODS allows you to encode all of the information:
  ```
  <dateCreated encoding="iso8601">1970-04-17</dateCreated>
  ```

Chapter 5 provides guidelines on which specific data fields are most desirable, how to structure the content of those fields, and how they may be encoded in XML format.

CHAPTER 2: DATA FIELD & STRUCTURE GUIDELINES

In RLG Cultural Materials each individual record belongs to a collection and can be discovered as part of a collection. A collection in the context of RLG Cultural Materials is any assembly of individual records that can be described as a whole, ranging from collections organized predominantly around work type ("Historic American Sheet Music"), creator ("Washington As It Was: Photographs By Theodor Horydczak"), or collector/collecting agency ("London School of Economics Pamphlet Collection").

The discovery mechanisms in RLG Cultural Materials are meant to bring together individual records that are only related through the user's search. If an individual descriptive record points to more than one digital surrogate (as a record for a manuscript might point at the images of its pages), these multi-part digital objects are represented in a METS viewer interface in RLG Cultural Materials. However, this does not mean you must submit METS objects. RLG creates METS objects from the descriptive data you submit.

Useful standards

There are many standards and specifications for organizing descriptive information about collections and individual items, all the result of significant consensus-building efforts by knowledgeable professionals in various curatorial and custodial professions. One of the goals of RLG's Cultural Materials Initiative is to integrate heterogeneous data meeting these different standards without requiring conformance to a single one.

Here are just some of the more common community standards we can work with (with the exception of VRA Core and CDWA, all have an officially sanctioned XML expression associated with them):

- Categories for the Description of Works of Art (CDWA) (for individual records)
- Dublin Core (DC) (for individual records)
- Encoded Archival Description (EAD) (for an organized collection of individual records)
- MARC 21 or UNIMARC (for individual records)
- Metadata Encoding and Transmission Standard (METS) (for digital objects)
- Metadata Object Description Schema (MODS) (for individual records)
- SPECTRUM (for collections of individual records)
- Text Encoding Initiative (TEI) (for individual records/digital objects)
- Visual Resources Association (VRA) Core Categories (for individual records)

For a more detailed and annotated list of these standards see Appendix A.

Descriptive strategies: collections, individual records, digital objects

An important first step in preparing descriptions is to determine the most appropriate descriptive approach and standard for the materials. This process of materials analysis considers how the materials are to be grouped and organized, and what aspects are to be described, how, and for what purposes. Going into even greater detail (and touching on themes to be explored in greater depth in Chapter 3), this analysis also investigates which data content rules, thesauri, or name authorities may be appropriate for the materials, and for which fields authority control has to be established. Data field and structure standards and their associated data formats range from entire structured collections of descriptive records to stand-alone individual descriptive records to so-called digital objects.

- **Collection description standards** organize individual descriptive records and relate them to one another, typically in a hierarchy. Traditionally, collection standards such as EAD have been used for archival materials that share the same provenance. However, EAD has also been used for a wider variety of materials, including museum objects, to present particular contextual arrangements of objects, or intentionally assembled collections (sometimes referred to as virtual or artificial collections). EAD allows you to describe materials both on an aggregate level for those characteristics shared by all the subordinate components, and on lower levels for those characteristics unique to individual items. Use of collection description standards is warranted in cases where the broader collection context is an important factor in understanding individual collection items.

- **Standards for individual descriptive records** usually rely on an external mechanism to bring related records together into collections, for example, in a database. Often, one of the fields in the record declares which larger unit or collection it belongs to, even if that larger unit is only the collecting institution itself. Individual records are best suited for describing materials that can stand on their own—that are meant to be understood and referenced as independent entities rather than as parts of a larger collection context.

- **Digital object standards** provide for describing both an item and its component parts, and organize the digital surrogates into a navigable whole. Digital objects are ideal for describing the internal structure of materials for which there may be multiple digital surrogates, for example, the many page images of a manuscript. METS is currently the most significant digital object standard, although TEI Guidelines for Text Encoding and Interchange can also be used for representing a digital object. METS may contain descriptive metadata for the item represented in a wide variety of individual record data structures, as long as they are expressed in a W3C (Worldwide Web Consortium) XML Schema.

General guidelines

Regardless of the strategy you use, following these general guidelines ensures that the data integrates effectively and performs well in RLG Cultural Materials:

- **Contribute the richest and most structured version of the descriptive data available.** For example, if you have both a set of detailed records organized into many fields or elements (such as a VRA Core record) and a set of Dublin Core records derived from the richer record, we can extract significantly more meaning from the original, detailed records than from the less detailed Dublin Core records.

- **Use fields consistently and in keeping with their official definitions.** Put the same type of information in the same field across all of the records in a collection. As many descriptive standards are designed to accommodate a wide range of practices, it is often useful to adhere to usage guidelines, which tailor the usage of a standard to a particular situation or implementation.

- **Avoid combining different types of data or repeating the same type of data in a single field.** Use separate fields, for example, when providing subjects pertaining to topics, places, and people.

- **Distinguish between entities associated with the creation of the materials described and those depicted in the materials.** For example, clearly distinguish between individuals and corporate bodies in their role as creators from individuals and corporate bodies in their role as subjects or topics.

- **Do not confuse description of the original materials and their digital surrogates.** It is hard to interpret the descriptive metadata if it contains descriptions of both items and digital surrogates and does not provide any way to distinguish between them.

- **Bear in mind that it might be necessary to have multiple fields for the same piece of information to support display, sorting, or searching**—for example, textual and numeric dates, or inverted names and names in direct order.

- **Consider adapting your existing records for submission to RLG Cultural Materials.** Often the information implied in a local context is lost upon transfer to a more general context such as RLG Cultural Materials. For example, it may be tacitly understood that all materials in a local database describing American postcards of the 19th century are postcards created in America during the 19th century, yet none of this information may be present in the records describing the materials.

It is highly recommended that you document your practices. Explanations of what fields and structures you have used, how, and for what purposes can greatly assist in the conversion of your contributed data. We appreciate receiving your local data dictionary or cataloging guidelines with your contribution.

CHAPTER 3: DATA CONTENT & VALUE GUIDELINES

Standards make data more portable and predictable across institutions and make it more likely that users will discover similar works from different institutions. For example, when several institutions select data values for creator names from the same standard, such as the Union List of Artist Names® (ULAN), users are more likely to retrieve all works by a given creator. Specific communities have published standards and guidelines for data content and values, however, institutions routinely use both published standards and locally developed authorities and guidelines. For example, an institution may use a published source, such as the ULAN, for creator names as well as a local name authority for names of regional publishers.

Useful standards

Specific communities often rely on different data content and value specifications. Libraries have traditionally looked to AACR2 for guidance, while the archival community is transitioning from Archives, Personal Papers, and Manuscripts (APPM) to Describing Archives: A Content Standard (DA:CS). For an annotated list of standards see Appendix B. The data content standard Cataloguing Cultural Objects (CCO) merits special attention in this context since it aims to bridge the library, museum, archival, and visual resources communities engaged in describing visual content. CCO presents a good starting point for investigating data content issues for the types of one-of-a-kind objects usually found in RLG Cultural Materials.

General guidelines

Observing a few guiding principles for data content and value will maximize the usefulness of your descriptions in RLG Cultural Materials. These principles are important to keep in mind when planning and documenting your local descriptive practices and when filling your information records with content.

- **Be consistent.** Entering data into fields consistently and predictably within and across your information records improves management of records at your institution and prepares your data for successful conversion, loading, and retrieval in RLG Cultural Materials. As the practical implementation of data content and value standards, your local cataloging rules should prescribe how to populate fields. For example, display names of known creators should be entered consistently (e.g., invert proper names for indexing) and creation dates should be formatted consistently (e.g., use ISO 8601 for indexing).

- **Be clear and specific.** Describe works clearly and at an appropriate level of specificity. Clear, well-chosen terms help users determine if a particular work suits their needs. Precise data values make your information records useful. Your local cataloging rules should address how to handle limited information about a work (e.g., unknown creator, unknown title) and how to clearly, consistently communicate that limitation to users.

- **Document your practice.** Your local cataloging guidelines should capture the choices your institution has made for describing works in your collection. Local guidelines help you ensure consistency across the different staff engaged in describing your collections and allow you to communicate your choices to others when you share your data.

CHAPTER 4: DATA FORMAT GUIDELINES

Data can be stored locally in any number of formats to meet the needs of the institution. The format you use to store data locally does not have to be the same as the format you use to transmit data to RLG or the final format RLG uses to manipulate the data. We prefer you transmit data in one of the formats described below because we have tools that can easily transform it for ingest into RLG Cultural Materials.

Ideal formats: XML & MARC

Our preferred data formats are XML and MARC. We also accept data in Standardized General Markup Language (SGML). Machine processing of the data, including transformation, is straightforward because explicit tagging makes data content easy to identify and manipulate, and character sets are identified. Multiple electronic tools are available for these formats and RLG is familiar and experienced with them.

We prefer XML that is encoded according to a standards-based XML Schema or document type definition (DTD). If you use a locally defined schema or DTD, submit the rules document along with the markup as well as a data dictionary explaining the meaning of each XML tag.

MARC is a standard transport mechanism RLG has worked with for many years. MARC data conventions may be embedded either in ISO 2709 format or as XML data, as in the Library of Congress's MARCXML Schema. You can use either the MARC 21 or UNIMARC field structures.

If desired, we can also harvest XML data from a server that conforms to the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH). While OAI serves Dublin Core as its mandatory default, the protocol may also transport other data formats such as MARCXML, MODS, or EAD.

Good formats: databases & spreadsheets

We can convert database files in Microsoft Access or FileMaker® Pro formats to XML using a variety of tools. A description of how the databases are structured helps us map to the RLG Cultural Materials schema.

You can use spreadsheets like Microsoft Excel as long as a detailed description of the cells' arrangement is available.

Acceptable formats: Comma- & tab-separated values

Easily manipulated, comma- or tab-separated value files are limited to a "flat file" format where data elements are delineated by the insertion of commas or tabs, with some other character used to separate records. Comma- or tab-separated values, while widely available as an export format from databases, may confound machine processing because of ambiguities about tabs and commas within the data and uncertainty about character sets. Usually, tab-separated is preferable to comma-separated because submissions are less likely

to contain tabs within the data. A description of the various values is required for mapping to the RLG Cultural Materials schema.

Other formats or upgrading

If you use a data format not listed above or you'd like to upgrade from a good format to an ideal one for submission, please make sure that your transformation preserves the maximum granularity of the data, meaning that all fields have been preserved and none of the data content that was previously entered in separate fields has been concatenated into a single field. Databases often have built-in export routines to a number of file formats, including XML. Usually, database export places data content within XML tags replicating the fieldname. For transformation of such a generic XML file to a document instance ruled by a standards-based schema, the programming language XSLT (Extensible Stylesheet Language Transformations) is a powerful tool.

A word of caution: our experience in the past shows that it usually is easier and more effective for RLG staff to transform databases to XML, rather than interpret the already transformed data. Especially if your database has not been built with a specific standard's export in mind, we'd prefer the database itself over extracted data. If you have questions about how to get from the data format in hand to an acceptable submission, your RLG content contact will be happy to help you. (To see your contact's name you can log on to the Cultural Materials Alliance Content Manager at culturalmaterials.rlg.org/collectron/.)

When transforming data it is essential to be aware of the character-encoding characteristics of the originating document, and the character-encoding capabilities of the transformation protocol.

A note on character sets

Character sets are the means for representing textual data so that it can be manipulated by computer. There is now a single universal character set, the Unicode™ Standard and its International Standard counterpart ISO/IEC 10646, for all of the world's scripts.

Before Unicode, many different character sets were used, each covering one or several scripts. These "legacy" character sets are still in use where an application has not been updated to use Unicode. ASCII, Latin-1, and Windows$^®$ code pages are examples of legacy character sets.

"ASCII" refers to a seven-bit character set, equivalent to International Standard ISO/IEC 646. "Latin-1" refers to International Standard ISO/IEC 8859-1, an eight-bit character set where the first 128 characters correspond to ASCII. Windows Code Page 1252 is a superset of Latin-1 (i.e., it includes additional characters), but is often incorrectly referred to as Latin-1.

Please provide as much information as possible about the character sets that have been used for your data. We can work with any published character set. If you are uncertain about the identity of the character set(s), please say so and we will thoroughly analyze your contribution before processing it. Character set issues may arise especially when you migrate data

from application to application or cut and paste text from various sources into a third document. The more you understand about how your various software applications handle characters sets, the fewer problems you will encounter.

For more information on character encoding in XML, MARC 21, and UNIMARC see Appendix D.

CHAPTER 5: CORE DESCRIPTIVE FIELDS

This chapter describes the data fields central to the functionality of RLG Cultural Materials and provides specific guidelines for their data content and values.

RLG Cultural Materials contains two types of descriptive records: collection records describe a group of related individual works; individual records each describe a single work, whether represented by a single digital surrogate or a multi-part digital object. The collection record is important because basic information such as collection title and collection contributor is inherited by and incorporated into associated individual records.

If you submit your collection through individual records, you can use the Cultural Materials Alliance Content Manager (culturalmaterials.rlg.org/collectron/) to provide contextual information on the collection as a whole, or consider sending us a record describing the entire collection in addition to the individual records. *All contributions to RLG Cultural Materials must include a collection record and individual records that contain a pointer to a digital surrogate or multiple surrogates.*

We have organized the possible data fields for individual records into three groups:

- Base-line: fields that should be included.
- Value-added: fields providing enhanced functionality in RLG Cultural Materials.
- Bonus: fields contributing to optimal functionality.

Since the overhead of preparing a collection description is minimal yet the functionality gains are significant, all collection record fields have been designated base-line. Of course, additional data fields are always welcome because RLG Cultural Materials is designed to take advantage of as much data as possible.

For each field, we provide a definition and a rationale based on five areas of functionality:

- Searching, which uses data that can be queried in RLG Cultural Materials.
- Browsing, which uses data that can be browsed in overview lists and results summaries.
- Refining, which uses data that can be queried to narrow a search result.
- Sorting, which uses data that can sort search results.
- Display, which uses data that appears in the brief list and full record display.

Searching, browsing, and refining are only as powerful as the data in RLG Cultural Materials. The central display for tapping into the riches of the database is called Results Overview, which shows three categories for the items in the result:

- What: work types, subjects (topical), collections, media types.
- Where: places depicted, places where works were created.
- Who: creators, contributors of works, depicted people/groups.

Users can trigger new searches from these categories, apply a date-range search, or refine the current result. Since the data driving Results Overview is crucial to finding and analyzing materials in RLG Cultural Materials, we've indicated which fields support it.

A Results Overview display in RLG Cultural Materials for a search on works created in the 19th century.

In the next section you will find an overview list of data fields for both a collection record and individual records. The entry for each data field includes guidance for data content and value. While the advice gives a solid starting point for investigating issues of data content and value, it will need to be supplemented with other applicable, more comprehensive specifications such as those listed in Appendix B. You will find every recommendation illustrated by a piece of XML markup in either EAD 2002 XML or MODS 3.0 XML. Our use of EAD and MODS is purely illustrative and does not constitute a preference for or endorsement of either standard for RLG Cultural Materials. For the complete examples of both markup standards see Appendix C.

Note: Most of the examples are based on collections and objects contributed to RLG Cultural Materials by the UC Berkeley Bancroft Library and the National Gallery of Canada, but they do not reflect actual cataloging from either institution. To illustrate pertinent issues, the sample records have been changed and augmented.

Summary of core fields

Collection record	**Your Mapping**

Base-line
- Collection Title* _____
- Collection Contributor* _____
- Collection Date Range _____
- Collection Narrative Description _____
- Collection Work Type* _____
- Collection Subjects (Topics, People, Places)* _____

Individual record	**Your Mapping**

Base-line
- Creator* _____
- Work Type* _____
- Work Title _____
- Creation Date _____
- Unique ID _____
- Pointer(s) to Surrogate(s) _____

Value-added
- Place of Creation* _____
- Subjects (Topics, People, Places)* _____
- Work Description _____

Bonus
- Sequence/Hierarchy of Surrogates (Multi-part work)
- Caption for Surrogates (Multi-part work) _____
- Measurement _____
- Materials and Techniques _____

* Supports Results Overview

A note for our international contributors: *For contributions in languages other than English, we require a translation of all the field labels you have used in describing your collections.* You can supply this information as, for example, a separate spreadsheet, mapping the original field label to its English equivalent. For the collection record we require translation of all data submitted. For individual records we require romanization or translation of titles and strongly encourage translation for work descriptions as well as the fields supporting the Results Overview (fields marked by an asterisk in the table above). Names and places can be romanized.

In RLG Cultural Materials we will display the data submitted in the native language (including non-Latin character sets) as well as the translated content. You also will be able to search the data using both the native language (including non-Latin character sets) as well as the translation.

Chapter 5: Core Descriptive Fields

Collection record

All of the data fields for collection records are considered base-line. The collection record provides an economical opportunity to characterize a large amount of material through a single descriptive record. As appropriate, collection-level data will be used to enrich item-level records during data conversion (see also Chapter 6). The Cultural Materials Alliance Content Manager (culturalmaterials.rlg.org/collectron/) captures all the information necessary for a base-line collection record; providing information through this Web site may substitute for submitting an actual collection record.

An example of a collection description in RLG Cultural Materials for the Robert B. Honeyman, Jr. Collection, submitted by the Bancroft Library, UC Berkeley.

Descriptive Metadata Guidelines for RLG Cultural Materials

COLLECTION RECORD, BASE-LINE FIELD

Collection Title
A descriptive title for a collection, striking a balance between concise, informative, and enticing. Collection titles may contain the name of the creator(s) or collector(s), the predominant form(s) of material, and the predominant topic(s). Used in searching, refining (with the What category in the Results Overview), and display functions. In some search results, the title provides the main clue about the content of a collection. The title should not refer to the project that generated the digital collection, but to the content of the collection itself.

Ideal: A title provided by the collection owner that is descriptive, for example, `Out of Time: the complete Grundy dinosaur paintings`

EAD 2002 XML
```
<unittitle label="Collection Name">Robert B. Honeyman,
Jr. Collection of Early Californian and Western
American Pictorial Material</unittitle>
```

MODS 3.0 XML
```
<titleInfo>
<title>Canadian Souvenir View Albums</title>
</titleInfo>
```

COLLECTION RECORD, BASE-LINE FIELD

Collection Contributor
The name of the institution or organization contributing the collection. Used in searching, refining, and sorting search results (with the Who category in the Results Overview). Use the name of your parent institution, not of an individual department.

Ideal: The official name of the contributing organization, for example, `Massachusetts Institute of Technology`

EAD 2002 XML
```
<publisher encodinganalog="260$b">The Bancroft Library,
University of California, Berkeley</publisher>
```

MODS 3.0 XML
```
<name type="corporate">
<namePart>National Gallery of Canada
</namePart>
<role>
<roleTerm type="code"
authority="marcrelator">ctb</roleTerm>
<roleTerm type="text"
authority="marcrelator">contributor</roleTerm>
</role>
</name>
```

Collection Date Range

A begin and end date representing the earliest and latest possible creation dates of items in the collection, not the period over which the collection was assembled, nor the period depicted, nor the time of creation of the surrogate images. Since date ranges are used both for searching (machine processing) and display (human processing), supplying both a numeric as well as text-based date range is ideal. For expressing uncertainty in ISO 8601 by using approximate dates see the RLG Best Practice Guidelines for Encoded Archival Description, p. 3 (www.rlg.org/en/page.php?Page_ID=450).

Ideal: Use ISO 8601 format (YYYY-MM-DD, divide date ranges by a forward slash) and a text-based date in a separate field, for example,
```
1970-04-17/1990-05-20; April 17, 1970 to May 20, 1990
1888/1988; 1888 to 1988
1756-11/1810-01; November 1756 to January 1810
```
Good: `1888 to 1988`

EAD 2002 XML
```
<unitdate type="inclusive" normal="1790/1930">1790 to
1930</unitdate>
```

MODS 3.0 XML
```
<originInfo>
<dateCreated>1880 to 1940</dateCreated>
<dateCreated encoding="iso8601"
point="start">1880</dateCreated>
<dateCreated encoding="iso8601"
point="end">1940</dateCreated>
</originInfo>
```

COLLECTION RECORD, BASE-LINE FIELD

Collection Narrative Description
A prose description of the content and context of the collection. Useful for keyword searching. It may include the size of the collection and indicate the diversity or uniformity of the materials in the collection.

EAD 2002 XML
```
<scopecontent>
<head>Collection Description</head>
<p> The Robert B. Honeyman Jr. Collection of Early
Californian and Western AmericanPictorial Material
comprises over 2300 items, with formats and media
ranging from original oil paintings, watercolors,
drawings, lithographs, engravings,
etchings,lettersheets, clipper cards, and ephemera, to
plates and spoons featuring western themes, and a fore-
edge painting. The collection focuses on pictorial
interpretations of the old West, with emphasis on the
early California and Gold Rush period.</p>
</scopecontent>
```

MODS 3.0 XML
```
<abstract>A collection of Canadian souvenir view
albums, published from the 1880s through the 1930s.
Having antecedents in the travel narratives,
topographical views and guidebooks that recorded and
illustrated the early periods of European exploration
and settlement of the country, souvenir view albums,
with their distinctive formats, a proliferation of
photomechanically reproduced images, and little or no
text, are peculiar to the late nineteenth and early
twentieth centuries.</abstract>
```

Collection Work Type

One or more terms, preferably from a controlled vocabulary source such as the AAT, describing the general characteristics of the items in the collection (e.g., their function or media type). Used in searching, browsing, refining search result sets (with the What category in the Results Overview), and display functions. AAT terms are preferred in order to standardize terminology across contributors. Please select the most specific terms and identify the source of your selected term in a separate field. For a list of source codes used with MODS consult the MODS Web site (www.loc.gov/standards/mods/).

Ideal: Terms drawn from the AAT. Select the most specific terms and use the plural form, for example, `albumen prints` (AAT), `playing cards` (AAT)

Good: Terms drawn from other published controlled vocabularies and cataloging rules, for example, `Offset Lithographs` (GMGPC)

EAD 2002 XML
```
<controlaccess>
<genreform source="gmgpc">Landscapes
(Representations)</genreform>
<genreform source="gmgpc">Oil Paintings</genreform>
<genreform source="gmgpc">Travel sketches</genreform>
<genreform source="aat">scientific
illustrations</genreform>
</controlaccess>
```

MODS 3.0 XML
```
<genre authority="aat">viewbooks</genre>
```

Descriptive Metadata Guidelines for RLG Cultural Materials

Collection Subjects (Topics, People, Places)

Terms, preferably from controlled vocabulary sources, used to describe the subject content represented in or by the collection. Used in searching and refining search results. Identify the source of your selected term in a separate field.

See Individual records: value-added fields, Subjects (page 30) for details.

EAD 2002 XML
```
<controlaccess>
<subject source="lctgm">Frontier &amp; pioneer
life</subject>
<persname role="subject" rules="aacr2">King, James,
1822-1856</persname>
<persname role="subject" source="lcnaf">Sutter, John
Augustus, 1826-1897</persname>
<persname role="subject" source="lcnaf">Fremont, John
Charles, 1813-1890</persname>
</controlaccess>
<controlaccess>
<geogname role="country of coverage"
source="tgn">United States</geogname>
<geogname role="region of coverage"
source="tgn">Western United States</geogname>
<geogname role="state of coverage"
source="tgn">California</geogname>
</controlaccess>
```

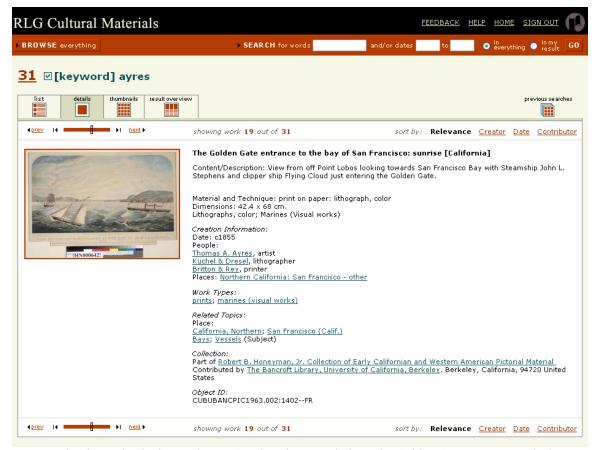**MODS 3.0 XML**
```
<subject authority="lctgm">
<topic>Photography</topic>
</subject>
<subject authority="lctgm">
<topic>Architecture</topic>
</subject>
<subject authority="lcsh">
<geographic>Canada</geographic>
</subject>
```

Individual record

For individual records the guidelines suggest six base-line fields, three value-added fields, and four bonus fields. Apart from describing the work, individual records also contain the all-important link between the record and the digital surrogate. Individual descriptive records may be accompanied by or be part of a complex digital object. If your collection contains materials that have one descriptive record but many digital surrogates (such as the pages of a manuscript), the fields Surrogate(s) (a base-line field), Caption for Surrogates (a bonus field), and Sequence for Surrogates (a bonus field) describe the process of communicating the relationship within a digital object to RLG. RLG Cultural Materials uses METS as its data structure for displaying multiple surrogates in a sequence or a hierarchy. We gladly accept METS objects, but we can also build METS objects from the minimal information described below.

An example of an individual record in RLG Cultural Materials for "The Golden Gate entrance to the bay of San Francisco: sunrise [California]," submitted by the Bancroft Library, UC Berkeley.

INDIVIDUAL RECORD, BASE-LINE FIELD

Creator   One or more name terms, preferably from a name authority source, for the person(s) or corporate body or bodies responsible for the creation or publication of the work. Used in searching, display, sorting, and refining search result sets (with the Who category in the Results Overview).

To optimize the use of names in RLG Cultural Materials:

▪ Select preferred names for individuals, corporate bodies, and cultural groups from name authorities or controlled vocabularies. If the name is not present in any published authority, create a form of the name based on available information and rules defined in content standards such as AACR2 and CCO.

▪ In the case of multiple creators, put each name in a separate field, repeating the field as necessary.

▪ To support browsing, invert personal names: last name, comma, first name (e.g. "Einstein, Albert") unless the parts of the name are already stored in separate fields. For display, the interface presents names in direct order. If possible, submit names in both forms. If only one is possible, please submit inverted names.

▪ Identify the source of the selected names (e.g., ULAN, LCNAF, local name authority) in a separate attribute or field.

▪ Differentiate consistently between names for persons, corporate bodies, and cultural groups.

When appropriate, clarify the role or relationship of the named persons or bodies to the work in a separate field; for example, "artist," "publisher." Use standard terms for roles and relationships whenever possible. Good starting points are the list of MARC relator terms at www.loc.gov/marc/sourcecode/relator /relatorlist.html and the ULAN role table linked to from www.getty.edu/research/conducting_research/vocabularies/ulan/

▪ Provide nationality and life dates to make identification easier.

▪ Provide a brief biography if possible.

Your local cataloging rules should provide direction for instances when no information is available about a creator. Local cataloging rules can enforce a consistency that supports more effective data transformation for adding collections to RLG Cultural Materials.

Creator
(cont'd)

Ideal: As many specific preferred names, in separate fields, as necessary to fully identify those responsible for making or publishing the work, their relationships to the work and brief biographical information, for example, `Anderloni, Faustino (Italian printmaker 1766-1847)` (person, ULAN); `Cluett & Sons (publisher)` (corporate body, local name authority)

Good: As many specific preferred names, in separate fields, as necessary to fully identify those responsible for making or publishing the work, for example, `Anderloni, Faustino`; `Cluett & Sons`

Acceptable: Those chiefly responsible for making the work, for example, `Anderloni, Faustino`

EAD 2002 XML
```
<origination>
<persname normal="Ayres, Thomas Almond" source="ulan"
role="artist">Ayres, Thomas Almond (ca. 1816-1858),
American painter</persname>
<corpname normal="Kuchel &amp; Dresel" source="lcnaf"
role="lithographer">Kuchel &amp; Dresel (active ca.
1853-ca. 1865), lithographer</corpname>
<corpname normal="Britton &amp; Rey" source="lcnaf"
role="printer">Britton &amp; Rey (active 1851-1902),
printer</corpname>
</origination>
```

MODS 3.0 XML
```
<name authority="local" type="personal">
<namePart>Glaser, Louis</namePart>
<displayForm>Louis Glaser</displayForm>
<role>
<roleTerm type="text"
authority="marcrelator">lithographer</roleTerm>
<roleTerm type="code"
authority="marcrelator">ltg</roleTerm>
</role>
</name>
```

Work Type One or more terms, preferably from a controlled vocabulary source such as the AAT, describing the general characteristics of the work (e.g., its function or media type). Used in searching, browsing, refining search results (with the What in the Results Overview), and display functions. AAT terms are preferred in order to facilitate browsing. Please select the most specific terms and use the plural form. Identify the source of your term in a separate field. For a list of source codes used with MODS consult the MODS Web site (www.loc.gov/standards/mods/).

Ideal: Terms drawn from the AAT. Select the most specific terms, for example, `albumen prints` (AAT), `playing cards` (AAT)
Good: Terms drawn from other published controlled vocabularies and cataloging rules, for example, `Panoramic Photographs` (GMGPC)
Acceptable: Term taken from a local authority.

EAD 2002 XML
```
<controlaccess>
<genreform source="aat">color lithographs</genreform>
<genreform source="aat">marines (visual
works)</genreform>
</controlaccess>
```

MODS 3.0 XML
```
<genre authority="aat">viewbooks</genre>
```

Work Title If known, the title for the work assigned by the creator or owner, or a descriptive title supplied by the collection contributor. Used in searching and display functions.

Ideal: A title assigned by the creator or owner or a descriptive title supplied by the cataloger, for example, `Moment in time` or `A laughing child on a bus`

EAD 2002 XML (assigned title)
```
<unittitle>The Golden Gate entrance to the bay of San
Francisco: sunrise</unittitle>
```

MODS 3.0 XML (supplied title)
```
<titleInfo>
<title>[Photolithograph illustrations from photographs
in Montreal]</title>
</titleInfo>
```

Creation Date   A date or date range for the creation or publication of the work. Used in searching and sorting functions. Since date ranges are used both for searching (machine processing) and display (human processing), supplying both a numeric as well as text-based date range is ideal. Dates can be a year (1888); a year and a month (1888-03); or a year, a month, and a day (1888-03-17). Text-based dates can be qualified to indicate uncertainty, using terms such as "circa," "before," and "after." For expressing uncertainty in ISO 8601 by using approximate dates see the RLG Best Practice Guidelines for Encoded Archival Description, p. 3 (www.rlg.org/en/page.php?Page_ID=450).

Ideal: Use ISO 8601 format (YYYY-MM-DD, divide date ranges by a forward slash) and a text-based date in a separate field, for example,
```
1969-08-24; August 24, 1969
1970-04-17/1990-05-20; April 17, 1970 to May 20, 1990
1756-11/1810-01; November 1756 to January 1810
1870/1899; 1870 to 1899
```
Good: `1870 to 1899`

EAD 2002 XML
```
<unitdate normal="1850/1860">circa 1855</unitdate>
```

MODS 3.0 XML
```
<dateIssued qualifier="approximate">1877</dateIssued>
<dateCreated encoding="iso8601"
point="start">1872</dateCreated>
<dateCreated encoding="iso8601"
point="end">1882</dateCreated>
```

Unique ID   The locally assigned alphanumeric identification number for the work, unique at least within the collection, preferably within the institution. Used in searching.

RLG uses identification numbers to manage, update, and/or remove individual records. Since RLG automatically assigns an institutional identifier to each record ID, no prefix is needed.

Ideal: Collection ID, plus the accession number or other locally unique or meaningful number, for example, `Poussin-67060077a`
Good: Collection ID, plus a running number unique to the collection, for example, `Poussin-1`

EAD 2002 XML
`<unitid>BANC PIC 1963.002:1402--FR</unitid>`

MODS 3.0 XML
`<identifier type="local">ONCA.b11310753</identifier>`

Pointer(s) to Surrogate(s)   A path plus a filename pointer linking the descriptive record to each digital surrogate of the work. The precise nature of the pointer depends on how you plan to submit your digital surrogates to RLG: it could be an HTTP URL if the surrogates are harvested from a Web site, an FTP URL if the surrogates are retrieved from an FTP site, or a directory path plus filename on a hard disk. RLG collects all surrogates and generates a new set of derivatives for RLG Cultural Materials. (For options for submitting your surrogates see Contributing to RLG Cultural Materials at www.rlg.org/en/page.php?Page_ID=523).

If more than one surrogate is associated with a record, simply listing the pointers in the correct sequence, or embedding sequence information into the filenames themselves (e.g., through a running number) allows RLG to build a display sequence. For an example of how to encode a sequence of images in MODS see the markup sample below. For additional examples of sequencing and other fields for a multi-part work see the bonus fields Sequence/Hierarchy for Surrogates (page 33) and Caption for Surrogates (page 33).

Pointer(s) to Surrogate(s) (cont'd)

EAD 2002 XML
```
<daogrp>
<daoloc
href="http://sunsite.berkeley.edu/FindingAids/dynaweb/c
alher/honeyman/figures/HN000642aA.jpg" role="hi-res"/>
</daogrp>
```

MODS 3.0 XML
```
<relatedItem type="constituent">
<titleInfo>
<title>Cover: Montreal</title>
</titleInfo>
<location>
<url>http://collections.ic.gc.ca/sva/qc/mon_01/images/i
mg_01.jpg</url>
</location>
</relatedItem>
<relatedItem type="constituent">
<titleInfo>
<title>Montreal and the St. Lawrence, Seen From Mount
Royal</title>
</titleInfo>
<location>
<url>http://collections.ic.gc.ca/sva/qc/mon_01/images/i
mg_02.jpg</url>
</location>
</relatedItem>
<relatedItem type="constituent">
<titleInfo>
<title>Notre Dame Cathedral and Place D'armes</title>
</titleInfo>
<location>
<url>http://collections.ic.gc.ca/sva/qc/mon_01/images/i
mg_03.jpg</url>
</location>
</relatedItem>
[…]
<relatedItem type="constituent">
<location>
<url>http://collections.ic.gc.ca/sva/qc/mon_01/images/i
mg_35.jpg</url>
</location>
</relatedItem>
```

Place of Creation Geographic name terms, preferably from a controlled vocabulary source like the Getty Thesaurus for Geographic Names® (TGN), indicating where the work (as opposed to the digital surrogate or the descriptive record) was created or published. Used in searching, browsing, and refining search result sets (with the Where category in the Results Overview). Works without an associated place name will be subsumed under the heading "unspecified," in the Results Overview.

To optimize the use of place names in RLG Cultural Materials:

- Put each part of the place name in a separate field so that the names of countries, states, and cities can be easily and unambiguously identified.

- If it is not possible to separate the parts of a place name, format the place name as a single string according to an applicable data content standard such as AACR2 or CCO.

- Identify the source of the selected terms (e.g., TGN, LCHS, local place name authority).

Ideal: United States [nation] California [state] Santa Clara county [county] Mountain View [inhabited place] (TGN)
Good: Mountain View (Santa Clara county, California, United States) (TGN); Mountain View (Calif.) (LCSH)
Acceptable: A term taken from a local authority.

EAD 2002 XML
```
<controlaccess>
<geogname role="country of origination"
source="tgn">United States</geogname>
<geogname role="state of origination"
source="tgn">California</geogname>
<geogname role="city of origination" source="tgn">San
Francisco</geogname>
</controlaccess>
```

MODS 3.0 XML
```
<place>
<placeTerm type="text">Montréal</placeTerm>
<placeTerm type="code"
authority="marccountry">quc</placeTerm>
</place>
```

Subjects (Topics, People, Places)

Terms, preferably from controlled vocabulary sources, used to describe the subject content represented in or by the work. Used in searching and refining search results.

We recommend contributing subject terms whenever possible because they provide rich access points. While different disciplines use different methods to describe subject content, for RLG Cultural Materials we recommend identifying the subject of a work by answering What?, Where?, and Who? For detailed guidelines see CCO or the Library of Congress Subject Headings discussion on applying subject terms. Local cataloging rules should provide direction for instances when subjects cannot be provided.

Cultural institutions may record subjects for things, topics, places, and persons in a single, repeated Subject field. In RLG Cultural Materials, dividing subject terms into the three categories of What, Where, and Who makes sorting and searching more effective and drives the performance of the Results Overview.

To optimize the use of your subject terms:

- Enter as many specific terms as necessary to capture the subject content represented in or by the work.

- Put each subject term in a separate, repeated data field.

- Select preferred terms from controlled vocabularies.

- Identify the source of the selected term (e.g., AAT, LCSH, ULAN, TGN, ICONCLASS, TGM, local authority).

- Differentiate subject terms among topical terms (for the What category), place names (for the Where category), or people or organizations (for the Who category).

- Be consistent; do not put all subject terms in one Subject field for some works and divide subject terms into the more specific fields for other works.

Subjects (Topics, People, Places) (cont'd)

Examples for Who: Follow the guidelines for Creator (page 23).

Examples for What category
Ideal: As many specific terms, in separate fields, as necessary to fully describe the topical subject or things depicted in or by the work, for example, `fans (costume accessories); festivals; fireworks (visual works); spectators`
Good: Selected terms, in separate fields, for the general subject content of the work, for example, `cityscapes; festivals; spectators`
Acceptable: One or more terms, in separate fields, for a key subject content of the work, for example, `festivals`

Examples for Where: Follow the guidelines for Place of Creation (page 29).

EAD 2002 XML (including Topics and Place)
```
<controlaccess>
<subject source="lctgm">Bays</subject>
<subject source="lctgm">Vessels</subject>
</controlaccess>
<controlaccess>
<geogname role="country of coverage"
source="tgn">United States</geogname>
<geogname role="state of coverage"
source="tgn">California</geogname>
<geogname role="city of coverage" source="tgn">San
Francisco</geogname>
<geogname role="site of coverage" source="tgn">Golden
Gate (channel)</geogname>
</controlaccess>
```

EAD 2002 XML (including Topics, People, and Place)
```
<controlaccess>
<subject source="lctgm">Criminals</subject>
<subject source="lctgm">Events</subject>
<persname role="subject" rules="aacr2">King, James,
1822-1856.</persname>
<persname role="subject" rules="aacr2">Casey, James
P.</persname>
<persname role="subject" rules="aacr2">Cora,
Charles</persname>
</controlaccess>
<controlaccess>
<geogname role="country of coverage"
source="tgn">United States</geogname>
<geogname role="state of coverage"
source="tgn">California</geogname>
<geogname role="city of coverage" source="tgn">San
Francisco</geogname>
</controlaccess>
```

Subjects (Topics, People, Places) (cont'd)

MODS 3.0 XML (including Topics and Place)
```
<subject authority="tgn">
<hierarchicalGeographic>
<country>Canada</country>
<province>Quebec</province>
<city>Montreal</city>
</hierarchicalGeographic>
</subject>
<subject authority="aat">
<topic>viewbooks</topic>
<topic>photography</topic>
</subject>
```

INDIVIDUAL RECORD, VALUE-ADDED FIELD

Work Description

A prose description of the content and context of a work—explaining, summarizing, or otherwise elaborating on the record for the work. Useful for keyword access.

Ideal: A description of the content and context of a work.
Good: A description of the content.

EAD 2002 XML
```
<scopecontent>
<p>View from off Point Lobos looking towards San
Francisco Bay with Steamship John L. Stephens and
clipper ship Flying Cloud just entering the Golden
Gate.</p>
</scopecontent>
```

MODS 3.0 XML
```
<note>This album, like many of the other Quebec albums,
was issued in the English language, a reminder of
English dominance of commerce in the province at the
time.</note>
```

Sequence/ Hierarchy of Surrogates (Multi-part Work)

The sequence and/or hierarchy in which surrogates should be arranged in the RLG Cultural Materials METS viewer. This information is often sufficiently implied through the order of filenames/pointers in a descriptive record, or through sequence information in the filename itself. Both Sequence/Hierarchy and Caption information can be submitted as part of a METS object, a separate spreadsheet (linked to from the appropriate descriptive record), or within the individual descriptive records. Especially if the structure of the multi-part work contains a nesting hierarchy, METS provides the most flexibility for encoding.

Ideal: Sequence/hierarchy information is conveyed in the structural map of a METS object.
Good: Sequence/hierarchy information is conveyed through the order of filenames given in a descriptive record, or through a spreadsheet linked from a descriptive record.
Acceptable: Sequence/hierarchy information is conveyed through sequence numbers within the filenames themselves.

Markup examples: See Caption for Surrogates (below)

Caption for Surrogates (Multi-part Work)

A concise label identifying the content of each individual surrogate. As titles for thumbnail images, used to facilitate browsing of a digital object. Caption information may be submitted in a separate spreadsheet, linked to from the appropriate descriptive record, or may be submitted in the individual descriptive records themselves.

Ideal: As appropriate, a label indicating a structural division inherent to the object (e.g. book-like objects, manuscripts), or a descriptive narrative label, for example, `Folio 10, Recto; Common Blackbird: Turdus merula`

EAD 2002 XML
Since the example EAD 2002 record used so far does not point to a multi-part work, this snippet of markup is derived from data submitted by the UC Berkeley Art Museum & Pacific Film Archive.

```
<daogrp>
<daoloc
href="http://www.bampfa.berkeley.edu/collections/bam/im
ages/moa2/bampfa_CT.20_1_2.jpg" role="hi-res"
title="album leaf 1"/>
```

Caption for Surrogates (Multi-part Work) (cont'd)	```<daoloc href="http://www.bampfa.berkeley.edu/collections/bam/images/moa2/bampfa_CT.20_3_2.jpg" role="hi-res" title="album leaf 3"/> […] </daogrp>```

MODS 3.0 XML

```
<relatedItem type="constituent">
<titleInfo>
<title>Cover: Montreal</title>
</titleInfo>
<location>
<url>http://collections.ic.gc.ca/sva/qc/mon_01/images/img_01.jpg</url>
</location>
</relatedItem>
<relatedItem type="constituent">
<titleInfo>
<title>Montreal and the St. Lawrence, Seen From Mount Royal</title>
</titleInfo>
<location>
<url>http://collections.ic.gc.ca/sva/qc/mon_01/images/img_02.jpg</url>
</location>
</relatedItem>
<relatedItem type="constituent">
<titleInfo>
<title>Notre Dame Cathedral and Place D'armes</title>
</titleInfo>
<location>
<url>http://collections.ic.gc.ca/sva/qc/mon_01/images/img_03.jpg</url>
</location>
</relatedItem>
  […]
<relatedItem type="constituent">
<location><url>http://collections.ic.gc.ca/sva/qc/mon_01/images/img_35.jpg</url>
</location>
</relatedItem>
```

METS XML

Since descriptive standards provide limited support for structuring multi-part works into hierarchies, they are often used in conjunction with METS. This XML example, derived from data submitted to RLG Cultural Materials by The Huntington Library, Art Collections and Botanical Gardens, shows parts of the structural map of a METS object with a nesting hierarchy. In this example the markup delineates that the manuscript "Ellesmere Chaucer" contains two chapters (or parts), "Knight's Tale" and "Prologue and Miller's Tale." These chapters are then represented by the individual pages

**Caption for
Surrogates
(Multi-part Work)**
(cont'd)

("Recto" and "Verso") of the manuscript. Note that the
<METS:fptr> tags point to an entry in the METS filegroup section
(not shown), which resolves the FILEID to a URL.

```
<METS:structMap>
      <METS:div ORDER="1" LABEL="Ellesmere Chaucer"
TYPE="Manuscript">
      <METS:div ORDER="1" LABEL="Knight&apos;s Tale"
TYPE="part">
            <METS:div ORDER="1" LABEL="Folio 10, Recto"
TYPE="folio">
                  <METS:fptr FILEID="FILE0506"/>
                  <METS:fptr FILEID="FILE0985"/>
                  <METS:fptr FILEID="FILE0027"/>
            </METS:div>
            <METS:div ORDER="2" LABEL="Folio 10, Verso"
TYPE="folio">
                  <METS:fptr FILEID="FILE0507"/>
                  <METS:fptr FILEID="FILE0986"/>
                  <METS:fptr FILEID="FILE0028"/>
            </METS:div>
            <METS:div ORDER="3" LABEL="Folio 11, Recto"
TYPE="folio">
                  <METS:fptr FILEID="FILE0508"/>
                  <METS:fptr FILEID="FILE0987"/>
                  <METS:fptr FILEID="FILE0029"/>
            </METS:div>
[…]
      </METS:div>
      <METS:div ORDER="2" LABEL="Prologue and
Miller&apos;s Tale" TYPE="part">
            <METS:div ORDER="1" LABEL="Folio 33, Verso"
TYPE="folio">
                  <METS:fptr FILEID="FILE0553"/>
                  <METS:fptr FILEID="FILE1032"/>
                  <METS:fptr FILEID="FILE0074"/>
            </METS:div>
            <METS:div ORDER="2" LABEL="Folio 34, Recto"
TYPE="folio">
                  <METS:fptr FILEID="FILE0554"/>
                  <METS:fptr FILEID="FILE1033"/>
                  <METS:fptr FILEID="FILE0075"/>
            </METS:div>
            <METS:div ORDER="3" LABEL="Folio 34, Verso"
TYPE="folio">
                  <METS:fptr FILEID="FILE0555"/>
                  <METS:fptr FILEID="FILE1034"/>
                  <METS:fptr FILEID="FILE0076"/>
            </METS:div>
[…]
      </METS:div>
[…]
      </METS:div>
</METS:structMap>
```

Measurement   The dimensions or size of a work or, for a time-based work, the length or running time. Used for display. Measurements allow users to determine the actual scale of the physical object on their screen, or the time they will have to devote to exploring it. Express physical measurements in metric units or inches, preferably both; unless indicated differently, the orientation implied is height by width by depth.

Ideal (dimension): Express in metric and inches, for example, `436.9 x 718 x 777 cm (14 feet 4 inches x 23 feet 6 1/2 inches x 25 feet 6 inches)`

Ideal (duration): Use text-based duration, for example, `1 year, 2 months, 3 days, 10 hours, 30 minutes, 27 seconds`

Good (dimension): Express in metric or inches, e.g., `40 x 60 cm`

EAD 2002 XML
```
<physdesc>
<dimensions unit="metric">42.4 x 68 cm</dimensions>
<dimensions unit="inches">17 inches x 27
inches</dimensions>
</physdesc>
```

MODS 3.0 XML
```
<physicalDescription>
<extent>9 x 13 cm</extent>
</physicalDescription>
```

Materials and Techniques   The substances an object consists of (materials) as well as the method of creation (techniques). Used for searching and display. Use free text or a controlled vocabulary such as the AAT or Thesaurus for Graphic Materials: TGM II, Genre and Physical Characteristic Terms (GMGPC).

Ideal: Express both materials and techniques using as many terms as needed to characterize this physical aspect of the work, for example, `pastels (crayons); chalk; charcoal paper` (AAT)

EAD 2002 XML
```
<physdesc>
<physfacet type="Materials and Techniques">print:
engraving, hand colored</physfacet>
</physdesc>
```

MODS 3.0 XML
```
<physicalDescription>
<form authority="aat">lithographs</form>
</physicalDescription>
```

Descriptive Metadata Guidelines for RLG Cultural Materials

CHAPTER 6: DATA CONVERSION FOR RLG CULTURAL MATERIALS

When RLG receives your contribution for loading into RLG Cultural Materials, our first steps involve analyzing the data, converting it to XML if necessary, and mapping it to the RLG Cultural Materials schema. The Data Resources group, part of RLG's Development department, is responsible for preparing the data for loading into all of RLG's databases. Each collection contributed to RLG Cultural Materials is assigned to one of the analysts, who thoroughly analyzes the data and designs and implements a conversion that is tailored to the collection:

- **Conversion to XML:** Data is contributed for RLG Cultural Materials in a wide variety of formats. If it is not already in XML, the first step is to convert it to this format. Various tools are used for this, including XMLSpy to convert Microsoft Access and FileMaker databases, and a MARC-to-XML converter developed by Virginia Tech and modified by RLG.

- **Creation of URL list for image collection:** The analyst uses the descriptive records provided by the contributor to produce a list of the URLs of the images. This list is then used by a program developed by RLG to collect images, either from the contributor's server or from an RLG server to which the images have been transferred. After the images are collected another RLG program processes the images to create high- and low-resolution derivatives that will go into RLG Cultural Materials.

- **Data analysis and mapping specifications:** A thorough analysis of the format and content of the data determines which data elements are present, how the contributor has used them, and what is the appropriate "home" for each data element in RLG Cultural Materials. Every collection is different, so few assumptions can be based on the format of the data or even on other data from the same contributor. Whenever possible, however, work done on other data in the same format or from the same contributor is reused. The analyst writes specifications that detail, for each data element in the RLG Cultural Materials database, which element(s) in contributed data will be mapped and what transformations are needed. Documentation provided by the contributor about their descriptive practice helps in this process.

- **Conversion development:** The analyst writes an XSLT (Extensible Stylesheet Language Transformations) that implements the mapping specifications and converts the contributed data into XML that conforms to the RLG Cultural Materials schema. This schema was designed specifically for loading data into the RLG Cultural Materials database. Presenting a collection in the best way in RLG Cultural Materials often involves transformations that are specific to the content, customized to each collection.

The analyst tests the stylesheet by examining the output, validating the output against the RLG Cultural Materials schema, and loading the converted data into a test RLG Cultural Materials database and examining the results. The test load is the first time that the images are linked to the descriptive data. After extensive testing and review, the collection is loaded into the production RLG Cultural Materials database.

Data transformations

An RLG dataloads analyst writes programs to convert data in batch. We do not see ourselves as editors of the data and we do not have the time or authority to go over each individual record. However, to provide the best possible presentation of data in RLG Cultural Materials, we often find ourselves doing data transformations that border on editing. Here are some examples of the types of transformations we have done:

Subjects: These three subject headings are all from the same collection, which was contributed in Dublin Core:

```
<dc:subject>Jackson County (Ind.)</dc:subject>
<dc:subject>Jefferson, Thomas, 1743-1826--Poetry</dc:subject>
<dc:subject>Land titles--Kentucky--Shelby County</dc:subject>
```

They are a place name, a personal name with a topical subdivision, and a topical heading with place subdivisions. In conversion, they must be categorized so that they will go into the different RLG Cultural Materials fields for place, name, and topical subjects.

Sometimes we can use pattern matching to roughly determine the type of subject heading, if the pattern holds for all data in a collection; for example, a comma followed by four digits might indicate a personal name with dates; data in parentheses could indicate a place. However, there are always many exceptions that do not follow the patterns.

We can also do a very content-specific coding of the stylesheet, such as, "If the field contains 'Kentucky,' it's a place." Content-specific coding means that if more works are added to a collection later, they must undergo the same level of analysis that was applied to the original data.

Our fallback position if we cannot determine the type of subject heading is either to categorize the subject headings manually (often assisted by automated pattern recognition) or to put all of the subjects from a collection into the What column of the Results Overview, where topical subjects belong.

Propagation of data from the collection or series level: Some contributions, particularly EAD, contain very little data at the item level. Data such as subject headings, creators, work types, and creation dates might occur only at the collection or series level, if at all. Sometimes this information is not in the contributed data at all, but only in the Content Manager, the tool used by contributors to describe and keep track of the collections they've submitted for RLG Cultural Materials.

To provide the best functionality in RLG Cultural Materials we want to have as many access points as possible at the item level. Some data fields, such as work type, are required in RLG Cultural Materials and if they are not present the analyst must assign a default value. If the analyst judges that collection- or series-level access points are appropriate to apply to all works in the collection or series, the propagation of these access points is built into the stylesheet.

Normalization of work types: To provide better access to work types in RLG Cultural Materials we normalize the terms provided by contributor to terms in the Objects Facet of the AAT. The term provided by the contributor is retained and displayed, and the normalized term is used in the Results Overview. For example, "stereographs," "daguerreotypes," and "negatives" are some of the many terms that are normalized to "photographs." Normalization of place names into a hierarchy is planned for the future.

Inversion of personal names: RLG Cultural Materials includes both a "sort form" and a "display form" for names. The sort form is browseable with the Who category in the Results Overview, and the display form appears in the Detail view. For personal names this means storing the name in inverted order (last name first) for sorting, and in direct order (first name first) for display. Birth and death dates are in separate fields in RLG Cultural Materials.

Here are four examples of data from a collection contributed in FileMaker Pro, all encoded in the Creator field. Multiple names are included in each field without any indication whether the content refers to a personal name or corporate name, an author, a lithographer, or an editor, for instance:

```
Boudreau, James C., Florence H. Fitch and Elmer A. Stephan
Snow, Bonnie E.; Froehlich, Hugo B.
Woolworth, Ainsworth & Co.
Perry, Walter Scott, et. al., eds.
```

This kind of data creates difficulties in conversion because:

- Personal and corporate names are not differentiated.
- Multiple names are combined into one field, with inconsistent separator characters (semi-colon, comma, or the word "and").
- Two of the personal names are in direct order, while the others are inverted.
- The creator's role (e.g., "eds.") is combined with the name.

Ideal data

While RLG can handle many kinds of data, we prefer standards-based XML. When you use a data value standard for a particular field, please clearly indicate its use. However you submit your data, please make sure that it is as granular and differentiated as possible, providing us with all the information you have gathered about a collection with each unit of information clearly delineated from the next.

APPENDIX A: DATA FIELD & STRUCTURE SPECIFICATIONS

Categories for the Description of Works of Art (CDWA)

The Categories for the Description of Works of Art (CDWA) are guidelines and a set of data fields for adequately describing works of art for the purposes of access and research. They articulate a hierarchical structure for the description of objects and images, which reveals the requirements and assumptions implicit in the discipline of art history. The CDWA specifies a set of data fields and a data structure, but not a data format (no official XML Schema or DTD). It can be implemented in a database or a locally developed XML Schema.

Current version (as of December 2004): CDWA 2.0

More information: www.getty.edu/research/conducting_research/standards/cdwa/

Use for: Developing element sets for collection management databases, interchange formats, or online union catalogs of art and architecture.

Recommendations: The data fields proposed by CDWA are comprehensive. The specification is probably most useful as a subset of elements carefully chosen for a specific purpose. For instance, one subset may be appropriate for developing an internal collection management system, while another may be more suitable for creating online public catalogs. The CDWA categories identified as "core" by the specification are a good starting point for creating subsets. The hierarchical structure allows information to be recorded at different levels of detail. Since many art museum projects use the CDWA, it may be useful to map local data to the CDWA as a preparatory exercise before submitting or exchanging data.

Tools: Since the CDWA does not have an official data format, there are no tools specific to CDWA. However, some commercial museum collection management systems are based on subsets of the CDWA elements, as are some project interchange formats, such as AMICO (Art Museum Image Consortium) and MOAC (Museums and the Online Archive of California).

Possible pitfalls: Attempting to use all of the possible elements for all purposes can be impractical and inappropriate.

Maintenance agents: J. Paul Getty Trust and College Art Association

Dublin Core (DC)

The Dublin Core Metadata Element Set, or Dublin Core (DC), is a set of 15 descriptive elements that provide a simple way to describe resources for more effective discovery and retrieval. Usually, DC is used to describe digital documents and information resources, but it has been adapted to other uses such as describing museum objects. Qualifiers can be applied to the 15 elements, either to identify a named encoding scheme or to refine (but not extend) the meaning of an element. DC has been expressed in RDF, XML, metatags in HTML documents, databases, etc.

Current version (as of December 2004): DC 1.1

More information: dublincore.org

Use for: Simple, brief description or minimal element set.

Recommendations: Contribute Dublin Core descriptions to RLG Cultural Materials only if it is not practical to contribute richer, more detailed descriptions.

If you use DC, we would prefer qualified Dublin Core, since it provides more detailed content designation; see dublincore.org/documents/dcmes-qualifiers/ for details.

Tools: A variety are available from dublincore.org/tools/.

Possible pitfalls: The most common pitfall is to confuse describing the object with describing the DC record itself. One simple DC record should describe one resource (document or object)—that is, for a museum object, the Title, Creator, and Date should be the title, creator, and date of the object. Similar information about the DC record itself should be omitted or contained outside the record.

Maintenance agents: Dublin Core Metadata Initiative

Encoded Archival Context (EAC)

Encoded Archival Context is a standard for the encoding of archival authority data using XML syntax.

Current version (as of December 2004): EAC beta version DTD and XML Schema

More information: www.iath.virginia.edu/eac/

Use for: Describing archival authority data, portions of which may be embedded either in EAC records or in EAD-encoded finding aids.

Recommendations: Use with data created according to the International Standard Archival Authority Record for Corporate Bodies, Persons, and Families (ISAAR (CPF)) or Describing Archives: A Content Standard.

Tools: A tag library is available.

Possible pitfalls: No currently available archival information systems support the separation of comprehensive authority data from descriptive information as conceptualized in EAC.

Maintenance agents: Ad Hoc Encoded Archival Context Working Group

Encoded Archival Description (EAD)

The Encoded Archival Description Document Type Definition (EAD DTD) is a standard for encoding archival finding aids using either SGML or XML. It provides a hierarchical descriptive framework that can be used to describe collections, groups, and items of materials. Version 2002 was finalized in December 2002. An XML Schema version is forthcoming.

Current version (as of December 2004): EAD 2002

More information: www.loc.gov/ead/

Use for: Primarily, describing archival collections, particularly those with a combination of collection, group, and item-level descriptions. Can also be used to describe collections or groups of museum objects.

Recommendations: Follow the RLG Best Practice Guidelines for Encoded Archival Description (www.rlg./en/page.php?Page_ID=450). While EAD has elements for relating digital surrogates of the original materials described, as with TEI (see below) it may be helpful to use METS alongside EAD in order to maintain greater control over the many associated digital files.

Tools: The EAD Cookbook is available at www.iath.virginia.edu/ead/cookbookhelp.html. The version of the text editor NoteTab at web.library.uiuc.edu/ahx/ead/tech/default.asp is a particularly useful, customizable, and inexpensive tool.

Possible pitfalls: EAD's flexibility can lead to inconsistent tagging. The same data may often be represented in different ways. For example, names associated with the materials described may be present in the <origination> and <unittitle> elements as well as in the <controlaccess> element.

Frequently a very coarse degree of granularity is employed, such as including the author, recipient, and date of a letter, with no encoding to distinguish each, in the <unittitle> element. Such encoding hinders extraction of data into a union database like RLG Cultural Materials.

Maintenance agents: The Network Development and MARC Standards Office at the Library of Congress and the Society of American Archivists (SAA)

MARC 21 Format & MARCXML

The MARC 21 formats published by the Library of Congress are standards for the representation and communication of bibliographic and related information in machine-readable form. Originally developed as a standard for the exchange of cataloging data, MARC has been modified over the years to provide a rich and flexible structure for the description of a wide range of bibliographic and related information types, including the description of digital resources. MARC is an implementation of the Format for Information Interchange (ISO 2709) and the Information Interchange Format (ANSI/NISO Z39.2) and can be exchanged via these formats or XML. MARC was first deployed by Library of Congress and other US libraries in 1968.

Current version (as of December 2004): MARC 21 through Update No. 4 (October 2003, updated annually)

More information:

For MARC 21: www.loc.gov/marc/bibliographic/.

For MARCXML: http://www.loc.gov/standards/marcxml/.

For the MARC family of standards: www.loc.gov/marc/.

Use for: A wide range of bibliographic and related information types.

Recommendations: MARC is a bibliographic format for books, continuing resources (serials), computer files, maps, music, sound recordings, motion pictures, visual materials, and archival collections. It is the most ubiquitous data format standard used in libraries and archives.

Tools: A number of MARC tools are available from commercial vendors along with new tools for working with MARCXML; see www.loc.gov/marc/marcservice.html. For specific MARCXML tools see www.loc.gov/standards/marcxml/index.html#stylesheets.

Possible pitfalls: The MARC formats are very structured and can provide catalogers with a daunting choice of data elements. While the MARC formats do not mandate the use of a particular set of cataloging rules, MARC is usually associated with standards for content, in particular the International Standard BibliographicDescription (ISBD) family of standards and Anglo-American Cataloguing Rules, 2nd edition (AACR2). The MARC formats are primarily used in libraries and archives.

Maintenance agents: The Network Development and MARC Standards Office at the Library of Congress, the Standards and Support Office at the Library and Archives Canada, and Bibliographic Standards and Systems at the British Library

Metadata Encoding & Transmission Standard (METS)

Metadata Encoding and Transmission Standard (METS) is an XML Schema for describing digital objects. This versatile standard can be used for data exchange, digital preservation, and online presentation. METS groups and organizes the metadata and the multiple files that compose a digital object into a single record. While it contains data elements for structural metadata and file information, METS relies on external standards such as EAD, MODS, MARC, and Dublin Core for descriptive metadata. A schema for METS Profiles has been developed for defining classes of instances to facilitate creation and interchange of METS files.

Current version (as of December 2004): METS Schema 1.3; METS Schema 1.4 under review.

More information: www.loc.gov/standards/mets/

Use for: Complex Digital Objects, that is, items that are represented by one descriptive record, but more than one digital surrogate.

Recommendations: Difficult to encode manually, METS files are best produced automatically as output from a database. When possible, create METS files according to the specifications included in a METS Profile.

Tools: Some tools for the creation and display of METS files can be found in the "Tools and Utilities" section of the METS Web page.

Possible pitfalls: As a flexible standard METS allows such a variety of encodings that it is easy to describe similar digital objects so differently that it impedes their processing and interchange. Furthermore, the intended function of a METS file (e.g., data exchange, digital preservation, or online presentation) may demand different data elements and degrees of granularity for describing the same digital object.

Maintenance agents: The Network Development and MARC Standards Office of the Library of Congress

Metadata Object Description Schema (MODS)

The Metadata Object Description Schema (MODS) is a bibliographic element set expressed as an XML Schema. It was designed by the Library of Congress in consultation with interested experts and includes a subset of the fields from MARC 21. MODS aims to strike a balance between simplicity and power, occupying a middle ground between basic Dublin Core and full-blown MARC 21 descriptions.

Current version (as of December 2004): MODS 3.0

More information: www.loc.gov/standards/mods/

Use for: A wide range of bibliographic and library-related applications.

Recommendations: MODS is designed to be a "human-readable" and simple version of MARCXML. MODS and MARCXML share common semantics, which enable users to map elements between schemas. MODS may be easier to implement because of its use of language-based tags. It is most likely to be used by libraries and archives that have already implemented MARC 21 or MARCXML.

Tools:
A MARCXML-to-MODS conversion stylesheet is available at www.loc.gov/standards/mods/v3/MARC21slim2MODS3.xsl.

A MODS-to-MARCXML conversion stylesheet is available at www.loc.gov/standards/marcxml/xslt/MODS2MARC21slim.xsl.

Possible pitfalls: Like MARCXML, MODS can be complex. Standard XML tools may be used; additional schema-specific tools (e.g., for input) are under development. As with any new standard, MODS is not in wide use yet, although its adoption is increasing.

Maintenance agents: The Network Development and MARC Standards Office at the Library of Congress

SPECTRUM: The UK Museum Documentation Standard

According to the mda Web site, SPECTRUM: The UK Museum Documentation Standard is a guide to "good practice for museum documentation, established in partnership with the museum community. It contains procedures for documenting objects and the processes they undergo, as well as identifying and describing the information which needs to be recorded to support the procedures." SPECTRUM comes out of the UK museum community and is maintained by mda. A CIMI project from 2000 to 2002 created an official SPECTRUM XML Schema. After the demise of CIMI, maintenance for the XML Schema reverted back to mda. As of the publication of these guidelines, SPECTRUM is undergoing a major revision, which will in turn affect the schema. This project is scheduled for completion by the end of the first quarter of 2005.

Current version (as of December 2004): SPECTRUM 2nd edition

More information: www.mda.org.uk/spectrum.htm

Use for: Description of museum objects.

Recommendations: Capturing extensive museum object descriptions and collections management information.

Tools: Because compatibility with SPECTRUM is a common requirement for UK museums, some commercially available collections management systems support the SPECTRUM Units of Information.

Possible pitfalls: SPECTRUM is in wide use in the UK, but American museums might find limited acceptance of SPECTRUM-formatted content by other museums or projects until it is more widely used in the US. Similarly, training and tools for SPECTRUM are currently hard to get in the US.

Maintenance agents: mda (formerly Museum Documentation Association)

Text Encoding Initiative Guidelines for Electronic Text Encoding & Interchange (TEI)

The Text Encoding Initiative Guidelines (TEI) provide data elements for the markup of a wide array of texts. It is expressed as an SGML or XML DTD, and in its upcoming P5 release will be available as a RelaxNG and XML Schema as well. TEI-Lite, a subset of the overall TEI DTD, has been used widely for electronic transcriptions of books, manuscripts, and other textual materials. Every TEI document contains a TEI Header for metadata describing the electronic transcription as well as its source and creation.

Current version (as of December 2004): TEI P4

More information: www.tei-c.org and sourceforge.net/projects/tei/

Use for: Transcriptions of textual materials. Can also include links to page images.

Recommendations: Consult the documentation for TEI-Lite (www.tei-c.org/Lite/) and for TEI-based projects. Useful guidelines have been made available by the University of Virginia Electronic Text Center (etext.lib.virginia.edu/tei/uvatei.html), the Women Writers Project (www.wwp.brown.edu/encoding/training/index.html), the Wright American Fiction Project (www.letrs.indiana.edu/web/w/wrightmrc/guidelines.html), and others listed under "Guides for Local Practice" at www.tei-c.org/Tutorials/index.html. For electronic transcriptions that include page images it might be beneficial to use TEI in conjunction with METS to organize and relate the digital image files.

Tools: A number of tools for creating TEI documents include TEI-Emacs (www.tei-c.org/Software/tei-emacs/) and <oXygen/> (www.oxygenxml.com/), which comes bundled with several TEI-specific features.

Possible pitfalls: Implementers of TEI may be tempted into "tag abuse," that is, employing an element for a purpose beyond its definition. For the TEI Header, it is important to keep in mind the distinction between the TEI instance (the electronic file itself) and the source from which it is derived.

Maintenance agents: The Text Encoding Initiative Consortium

UNIMARC Format & XML UNIMARC

UNIMARC (or "Universal MARC") formats are standards for the representation and communication of bibliographic and related information in machine-readable form. Origi nally developed in 1977 by IFLA (International Federation of Library Associations and Institutions), UNIMARC was created to facilitate the international exchange of bibliographic data in machine-readable form between national bibliographic agencies. Designed to provide a common record conversion format for the diverse national MARC formats, UNIMARC is an implementation of the Format for Information Exchange (ISO 2709).

Current version (as of December 2004): UNIMARC manual: bibliographic format/IFLA UBCIM. 2nd ed. München: K.G. Saur, 1994. (Update 1, 1996; Update 2, 1998; Update 3, 2000; Update 4, 2002). UNIMARC manual: bibliographic format (through Update 3, 2000): www.ifla.org/VI/3/p1996-1/sec-uni.htm

More information:
UNIMARC standard: www.ifla.org/VI/3/p1996-1/unimarc.htm

XML for UNIMARC, Draft UNIMARC/XML package: www.bookmarc.pt/unimarc/

UNIMARC to MARC 21 conversion: www.loc.gov/marc/unimarctomarc21.html

Use for: A wide range of bibliographic and related information types.

Recommendations: MARC is a bibliographic format for books, continuing resources (serials), computer files, maps, music, sound recordings, motion pictures, visual materials, and archival collections.

Tools: A number of UNIMARC tools are available from commercial vendors along with new tools for working with UNIMARC/XML.

Possible pitfalls: The UNIMARC formats are very structured and can provide catalogers with a daunting choice of data elements mapped from the various national MARC formats. While the UNIMARC formats do not mandate the use of a particular set of cataloging rules, UNIMARC is usually associated with standards for content, in particular ISBD. The MARC formats are primarily used in libraries and archives.

Maintenance agents: The IFLA UNIMARC Core Activity has been hosted by the Biblioteca Nacional de Portugal since 2003, with maintenance and promotion of UNIMARC controlled by IFLA's Permanent UNIMARC Committee (PUC), founded in 1991. UNIMARC was formerly part of IFLA's Universal Bibliographic Control and International MARC (UBCIM) Core Activity, hosted by the British Library (1973-1989) and Die Deutsche Bibliothek (1990–2003).

Visual Resources Association (VRA) Core Categories

The VRA Core Categories, Version 3.0 consist of a single element set that can be applied as many times as necessary to create records to describe works of visual culture, as well as the images that document them, in a shared environment. In Version 3.0 the two distinct element sets from Version 2.0 (Work and Visual Document) have been conflated into a single element set (the distinction between Work and Visual Document is now made by assigning a value of either "work" or "image" to the new Record Type element). Qualifiers to provide more detailed semantics are also suggested for some elements.

Current version (as of December 2004): VRA 3.0

More information: www.vraweb.org/vracore3.htm

Use for: Descriptions of visual resource materials.

Recommendations: As an element set the VRA Core strikes a good balance between simplicity and descriptive richness.

Tools: The Data Standards Committee of VRA is currently developing an XML Schema for VRA Core.

Possible pitfalls: The VRA Core is a subset of the larger element set Categories for the Description of Works of Art, and lacks some elements necessary for describing primary source collections, such as ownership history/provenance information.

Maintenance agents: Visual Resources Association Data Standards Committee

APPENDIX B: DATA CONTENT & VALUE SPECIFICATIONS

With thanks, this list is based on a document created by the Getty Vocabulary Program. For a list of source codes for referencing data content and value standards see the MODS Web site (www.loc.gov/standards/mods/).

Anglo-American Cataloguing Rules, Second Edition, 2002 Revision: 2004 (AACR2)
A content standard designed for use in constructing catalogs and other lists of materials held in general libraries. The rules cover description of library materials, formulation of names, and provisions for access points. The current text is the Second Edition, 2002 Revision: 2004 which incorporates all changes approved by the Joint Steering Committee for Revision of AACR (JSC) through February 2004.

Anglo-American Cataloguing Rules. 2nd ed., 2002 revision: 2004. Chicago: American Library Association, 2004.

Archives, Personal Papers, and Manuscripts: A Cataloging Manual for Archival Repositories, Historical Societies, and Manuscript Libraries (2nd ed.) (APPM)
Superseded by DA:CS. A content standard for description and cataloging of archival materials. Focuses on collection-level description and access points that can be integrated into bibliographic catalogs constructed using Anglo-American Cataloguing Rules, 2nd edition (AACR2). Contains several sample descriptions, MARC-coded versions for all examples given in the text, and tables relating the manual's descriptive elements and MARC fields.

Hensen, Steven L. *Archives, personal papers, and manuscripts: a cataloging manual for archival repositories, historical societies, and manuscript libraries*. Chicago: Society of American Archivists, 1989.

Art & Architecture Thesaurus® (AAT)
www.getty.edu/research/conducting_research/vocabularies/aat/
A structured vocabulary that contains terminology and other information related to art, architecture, and related disciplines. Updated monthly.

Cataloguing Cultural Objects (CCO)
www.vraweb.org/CCOweb/
Guidelines for selecting, ordering, and formatting data used to populate catalog records. CCO is designed to promote good descriptive cataloging, shared documentation, and enhanced end-user access.

Categories for the Description of Works of Art (CDWA)
www.getty.edu/research/conducting_research/standards/cdwa/
A metadata schema that describes a set of categories and recommendations that may be used to design information systems and to do cataloging for art, architecture, and related materials, including archaeological and archival materials. Illustrated cataloging examples are given at www.getty.edu/research/conducting_research/standards/cdwa/3_cataloging_examples/.

Appendix B: Data Content & Value Specifications

Describing Archives: A Content Standard (DACS)
Supersedes APPM. Designed to facilitate consistent, appropriate, and self-explanatory description of archival materials and creators of archival materials.

Describing Archives: A Content Standard. Chicago: Society of American Archivists. 2004.

Getty Thesaurus of Geographic Names® (TGN)
www.getty.edu/research/conducting_research/vocabularies/tgn/
A structured vocabulary containing names and other information about geographic places, focusing on but not limited to places important to art, architecture, and related disciplines. Contains more than 1 million records. Updated monthly.

Graphic Materials
www.loc.gov/rr/print/gm/graphmat.html
A set of rules for the descriptive cataloging of prints, photographs, and other pictorial materials. Does not include rules for subject cataloging and classification. Compatible with Anglo-American Cataloguing Rules, 2nd edition (AACR2).

Betz, Elisabeth. *Graphic Materials: Rules for Describing Original Items and Historical Collections*. Washington, DC: Library of Congress, 1982.

The 1997 update (with optional additional rules and examples in MARC content designation) is available for downloading at the Graphic Materials Web site. The 1996 update text is online at www.itsmarc.com/crs/grph0199.htm.

International Standard Archival Authority Record for Corporate Bodies, Persons, and Families (ISAAR (CPF))
www.ica.org
The International Standard Archival Authority Record is a set of 26 elements used to identify the authoritative and variant forms of names associated with archival materials. Other elements contain additional contextual information about these entities such as biographical or administrative histories and relationships. A third set of elements link these entities to additional resources by or about them.

ICONCLASS
www.iconclass.nl
A classification system for iconographic research and documentation of images. ICONCLASS is a collection of ready-made definitions of objects, persons, events, situations, and abstract ideas that can be the represented in the visual arts. The system is organized in ten broad divisions within which subjects are ordered hierarchically. In the Web version, the narrative descriptions are associated with keywords that enhance subject access.

Also available in print: Waal, Henri van de. *ICONCLASS: an Iconographic Classification System*. 9 vols., completed and edited by L.D. Couprie with E. Tholen, G. Vellekoop. Amsterdam; New York: North-Holland Pub. Co., 1973–1983.

LC/NACO Authority File

authorities.loc.gov

A collection of authority records for names (personal, corporate, meeting/conference, geographic jurisdictions) and titles (uniform titles, series) established by the Library of Congress and cooperating libraries from around the world via NACO, the name authority component of the Program for Cooperative Cataloging. Available free of charge for individual record searching through the Library of Congress Web site; the full file is also made available for a fee in MARC 21 or MARCXML formats by the Library of Congress Cataloging Distribution Service and other vendors. The LC/NACO Authority File contains over 6 million records and grows by over a quarter of million records each year.

Library of Congress Subject Headings (LCSH)

authorities.loc.gov

LCSH is the largest general subject indexing vocabulary in the English language, covering all subject areas, developed to provide subject access points for use in bibliographic records for all types of library materials. Contributions to LCSH come from Library of Congress cataloging operations as well as contributions from libraries around the world via SACO, the subject authority component of the Program for Cooperative Cataloging. LCSH is not a true thesaurus in the sense that it is not a comprehensive list of all valid subject headings; rather LCSH combines authorities (now five volumes in their printed form) and a number of auxiliary aids. Most important is the four-volume *Subject Cataloging Manual: Subject Headings*, a manual detailing requirements for creating headings that are not directly established in the authority file, and *Free-floating Subdivisions: An Alphabetical Index*.

LCSH is available in an annual print edition (2005 is the 28th edition), and the authority records on which the print version is based are available free of charge for individual record searching through the Library of Congress Web site; the full file (almost 300,000 records) is also made available for a fee in MARC 21 or MARCXML formats by the Library of Congress Cataloging Distribution Service and, and also available on the Web via Classification Web (www.loc.gov/cds/classweb/).

Moving Image Materials: Genre Terms.

A source of terms for use in cataloging or indexing motion pictures.

Yee, Martha M. comp. *Moving image materials: genre terms*. 1st ed. Washington, DC: Cataloging Distribution Service, Library of Congress, 1988.

Rules for Archival Description (RAD)

www.cdncouncilarchives.ca/archdesrules.html

RAD2 Draft: www.cdncouncilarchives.ca/rad2.html

RAD/EAD Crosswalk (Canadian Committee on Archival Description, revised November 2003: www.cdncouncilarchives.ca/CrosswalkEN_Nov03.pdf

Designed to facilitate consistent, appropriate, and self-explanatory description of archival materials and creators of archival materials. Prepared under the direction the Bureau of Canadian Archivists, representing: L'Association des archivistes du Québec and The Association of Canadian Archivists. RAD2 (available in draft) will eventually supersede RAD (version 1990).

Subject Index for the Visual Arts
A structured vocabulary used for subject access to materials in the Print Room of the Victoria & Albert Museum, London.

Glass, Elizabeth, comp. *A Subject Index for the Visual Arts*. London: H. M. Stationery Office, 1969.

Thesaurus for Graphic Materials (TGM)
www.loc.gov/rr/print/tgm1/
www.loc.gov/rr/print/tgm2/
A structured vocabulary containing terms that describe both the subjects and the work types of visual materials. Nearly 7,000 subject terms provide vocabulary to index activities, objects, people, events, and facilities shown in pictures as well as broad concepts such as agriculture and civil rights. More than 600 genre and physical characteristic terms cover formats and media for drawings, photographs, prints, and pictorial ephemera.

Library of Congress, Prints and Photographs Division. *Thesaurus for Graphic Materials: Subject Terms* (TGM I) and *Thesaurus for Graphic Materials: Genre and Physical Characteristic Terms* (TGM II). 1995 -present; updated monthly.

Union List of Artist Names® (ULAN)
www.getty.edu/research/conducting_research/vocabularies/ulan/
A structured vocabulary containing names and other information on artists, architects, and other creators of art and architecture. Updated monthly.

APPENDIX C: DATA FORMAT XML EXAMPLES

The markup examples below combine examples from Chapter 5 into a complete and valid XML document instance.

To view and analyze the examples more easily in your XML editor, you can download them as separate XML files at: www.rlg.org/en/page.php?Page_ID=214.

Our use of EAD and MODS is purely illustrative and does not constitute preference for or endorsement of either standard for RLG Cultural Materials. These examples only show one possible way to use the units of information provided by the EAD and MODS to encode the data fields and data values discussed in these guidelines. Real-life description has to grapple with additional complexities that are out the scope of this document.

These examples are based on collections and objects contributed to RLG Cultural Materials by the UC Berkeley Bancroft Library and the National Gallery of Canada, but they do not reflect actual cataloging from either institution. To illustrate pertinent issues, the sample records have been changed and augmented.

Example 1: EAD 2002 XML

```xml
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE ead PUBLIC "+//ISBN 1-931666-00-8//DTD ead.dtd (Encoded Archival Description (EAD) Version 2002)//EN" "http://www.aip.org/history/ead/dtd/2002/ead.dtd">
<ead>
   <eadheader>
      <eadid publicid="PUBLIC &quot;-//University of California, Berkeley::Bancroft Library//TEXT (US::CU-BANC::BANC PIC 1963.002:0001-1886::Robert B. Honeyman, Jr. Collection of Early Californian and Western American Pictorial Material)//EN&quot; &quot;honeyman.xml&quot;" countrycode="us" mainagencycode="CU">PUBLIC "-//University of California, Berkeley::Bancroft Library//TEXT (US::CU-BANC::BANC PIC 1963.002:0001-1886::Robert B. Honeyman, Jr. Collection of Early Californian and Western American Pictorial Material)//EN" "honeyman.xml"</eadid>
      <filedesc>
         <titlestmt>
            <titleproper>Robert B. Honeyman, Jr. Collection of Early Californian and Western American Pictorial Material</titleproper>
         </titlestmt>
         <publicationstmt>
            <publisher>University of California, Berkeley</publisher>
         </publicationstmt>
      </filedesc>
   </eadheader>
   <archdesc level="collection">
      <did>
         <head>Descriptive Summary</head>
         <unittitle label="Collection Name">Robert B. Honeyman, Jr. Collection of Early Californian and Western American Pictorial Material</unittitle>
         <unitdate normal="1790/1930">1790 to 1930</unitdate>
      </did>
      <scopecontent>
         <head>Collection Description</head>
         <p>
```

Appendix C: Data Format XML Examples

The Robert B. Honeyman Jr. Collection of Early Californian and Western American Pictorial Material is comprised of over 2300 items, with formats and media ranging from original oil paintings, watercolors, drawings, lithographs, engravings, etchings, lettersheets, clipper cards, and ephemera, to plates and spoons featuring western themes, and a fore-edge painting. The collection focuses on pictorial interpretations of the old West, with emphasis on the early California and Gold Rush periods…</p>
        </scopecontent>
        <controlaccess>
           <genreform source="gmgpc">Landscapes</genreform>
           <genreform source="gmgpc">Photographs</genreform>
           <genreform source="gmgpc">Travel sketches</genreform>
        </controlaccess>
        <controlaccess>
           <subject source="lctgm">frontier & pioneer life</subject>
           <geogname role="subject" source="tgn">United States -- California</geogname>
           <persname role="subject" source="lcnaf">King, James, 1822-1856.</persname>
           <persname role="subject" source="lcnaf">Sutter, John Augustus, 1826-1897.</persname>
           <persname role="subject" source="lcnaf">Fremont, John Charles, 1813-1890.</persname>
        </controlaccess>
        <dsc type="in-depth">
           <head>Container List</head>
           <c01 level="series">
              <did>
                <unittitle>Individual items by geographical location</unittitle>
              </did>
              <c02 level="subseries">
                <did>
                  <unittitle>Northern California: San Francisco</unittitle>
                </did>
                <c03 level="item">
                  <did>
                    <daogrp>
                      <daoloc
href="http://sunsite.berkeley.edu/FindingAids/dynaweb/calher/honeyman/figures/HN000642aA.jpg"
role="hi-res"/>
                    </daogrp>
                    <unittitle>The Golden Gate entrance to the bay of San Francisco</unittitle>
                    <unitdate normal="1850/1860">circa 1855</unitdate>
                    <physdesc>
                      <dimensions unit="metric">42.4 x 68 cm</dimensions>
                      <dimensions unit="inches">1 foot 5 inches x 2 feet 3 inches</dimensions>
                      <physfacet type="Materials and Techniques">print on paper: engraving, hand colored</physfacet>
                    </physdesc>
                    <origination>
                      <persname normal="Ayres, Thomas Almond" source="ulan"
role="artist">Ayres, Thomas Almond (ca. 1816-1858), American painter</persname>
                      <corpname normal="Kuchel & Dresel" source="lcnaf"
role="lithographer">Kuchel & Dresel (active ca. 1853-ca. 1865), lithographer</corpname>
                      <corpname normal="Britton & Rey" source="lcnaf" role="printer">Britton
& Rey (active 1851-1902), printer</corpname>
                    </origination>
                    <unitid>BANC PIC 1963.002:1402--FR</unitid>
                  </did>
                  <scopecontent>
                    <p>View from off Point Lobos looking towards San Francisco Bay with Steamship
John L. Stephens and clipper ship Flying Cloud just entering the Golden Gate.</p>
                  </scopecontent>
                  <controlaccess>
                    <genreform source="aat">color lithograph</genreform>

        Descriptive Metadata Guidelines for RLG Cultural Materials

```
                    <genreform source="aat">marine (visual work)</genreform>
                </controlaccess>
                <controlaccess>
                    <subject source="lctgm">Bays</subject>
                    <subject source="lctgm">Vessels</subject>
                    <geogname role="subject" source="tgn">United States -- California -- San
Francisco -- Golden Gate (channel)</geogname>
                    <geogname role="place of creation" source="tgn">United States -- California --
San Francisco</geogname>
                </controlaccess>
            </c03>
        </c02>
    </c01>
</dsc>
</archdesc>
</ead>
```

Example 2: MODS XML

Collection record

```
<?xml version="1.0" encoding="UTF-8"?>
<mods xmlns="http://www.loc.gov/mods/v3" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns:xlink="http://www.w3.org/1999/xlink" version="3.0" xsi:schemaLocation="http://www.loc.gov/mods/v3
http://www.loc.gov/standards/mods/v3/mods-3-0.xsd">
    <titleInfo>
        <title>Canadian Souvenir View Albums</title>
    </titleInfo>
    <name type="corporate">
        <namePart>National Gallery of Canada
</namePart>
        <role>
            <roleTerm type="code" authority="marcrelator">ctb</roleTerm>
            <roleTerm type="text" authority="marcrelator">contributor</roleTerm>
        </role>
    </name>
    <typeOfResource collection="yes">text</typeOfResource>
    <genre authority="lctgm">viewbooks</genre>
    <originInfo>
        <dateCreated>1880 to 1940</dateCreated>
        <dateCreated encoding="iso8601" point="start">1880</dateCreated>
        <dateCreated encoding="iso8601" point="end">1940</dateCreated>
    </originInfo>
    <abstract>A collection of Canadian souvenir view albums, published from the 1880s through the 1930s.
Having antecedents in the travel narratives, topographical views and guidebooks that recorded and
illustrated the early periods of European exploration and settlement of the country, souvenir view albums,
with their distinctive formats, a proliferation of photomechanically reproduced images, and little or no text,
are peculiar to the late nineteenth and early twentieth centuries.</abstract>
    <subject authority="lctgm">
        <topic>Photography</topic>
    </subject>
    <subject authority="lctgm">
        <topic>Architecture</topic>
    </subject>
    <subject authority="lcsh">
        <geographic>Canada</geographic>
    </subject>
</mods>
```

Appendix C: Data Format XML Examples

Individual item record

```xml
<?xml version="1.0" encoding="UTF-8"?>
<mods xmlns="http://www.loc.gov/mods/v3" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns:xlink="http://www.w3.org/1999/xlink" version="3.0" xsi:schemaLocation="http://www.loc.gov/mods/v3
http://www.loc.gov/standards/mods/v3/mods-3-0.xsd">
    <titleInfo>
        <title>[Photolithograph illustrations from photographs in Montreal]</title>
    </titleInfo>
    <name authority="local" type="personal">
        <namePart>Glaser, Louis</namePart>
        <displayForm>Louis Glaser</displayForm>
        <role>
            <roleTerm type="text" authority="marcrelator">lithographer</roleTerm>
            <roleTerm type="code" authority="marcrelator">ltg</roleTerm>
        </role>
    </name>
    <genre authority="aat">viewbook</genre>
    <originInfo>
        <dateIssued qualifier="approximate">1877</dateIssued>
        <dateCreated encoding="iso8601" point="start">1872</dateCreated>
        <dateCreated encoding="iso8601" point="end">1882</dateCreated>
        <place>
            <placeTerm type="text">Canada -- Quebec -- Montreal</placeTerm>
            <placeTerm type="code" authority="marccountry">quc</placeTerm>
        </place>
    </originInfo>
    <physicalDescription>
        <extent>9 x 13 cm</extent>
        <form authority="aat">lithograph</form>
    </physicalDescription>
    <note>This album, like many of the other Quebec albums, was issued in the English language, a
reminder of English dominance of commerce in the province at the time.</note>
    <subject authority="tgn">
        <hierarchicalGeographic>
            <country>Canada</country>
            <province>Quebec</province>
            <city>Montreal</city>
        </hierarchicalGeographic>
    </subject>
    <subject authority="aat">
        <topic>viewbook</topic>
        <topic>photography</topic>
    </subject>
    <relatedItem type="constituent">
        <titleInfo>
            <title>Cover: Montreal</title>
        </titleInfo>
        <location>
            <url>http://collections.ic.gc.ca/sva/qc/mon_01/images/img_01.jpg</url>
        </location>
    </relatedItem>
    <relatedItem type="constituent">
        <titleInfo>
            <title>Montreal and the St. Lawrence, Seen From Mount Royal</title>
        </titleInfo>
        <location>
            <url>http://collections.ic.gc.ca/sva/qc/mon_01/images/img_02.jpg</url>
        </location>
```

Descriptive Metadata Guidelines for RLG Cultural Materials

```
        </relatedItem>
        <relatedItem type="constituent">
            <titleInfo>
                <title>Notre Dame Cathedral and Place D'armes</title>
            </titleInfo>
            <location>
                <url>http://collections.ic.gc.ca/sva/qc/mon_01/images/img_03.jpg</url>
            </location>
        </relatedItem>
        <relatedItem type="constituent">
            <titleInfo>
                <title>Post Office, Bank of Montreal, Mount Royal Park</title>
            </titleInfo>
            <location>
                <url>http://collections.ic.gc.ca/sva/qc/mon_01/images/img_04.jpg</url>
            </location>
        </relatedItem>
        <relatedItem type="constituent">
            <titleInfo>
                <title>Victoria Square, Y.M.C.A. Building, St. Andrews Church</title>
            </titleInfo>
            <location>
                <url>http://collections.ic.gc.ca/sva/qc/mon_01/images/img_05.jpg</url>
            </location>
        </relatedItem>
        <relatedItem type="constituent">
            <titleInfo>
                <title>James Street, Molson's Bank, Mechanics Hall, Post Office</title>
            </titleInfo>
            <location>
                <url>http://collections.ic.gc.ca/sva/qc/mon_01/images/img_06.jpg</url>
            </location>
        </relatedItem>
        <relatedItem type="constituent">
            <titleInfo>
                <title>Notre Dame Street, looking west, Nelson's Monument, Court House</title>
            </titleInfo>
            <location>
                <url>http://collections.ic.gc.ca/sva/qc/mon_01/images/img_07.jpg</url>
            </location>
        </relatedItem>
        <relatedItem type="constituent">
            <titleInfo>
                <title>New City Hall</title>
            </titleInfo>
            <location>
                <url>http://collections.ic.gc.ca/sva/qc/mon_01/images/img_08.jpg</url>
            </location>
        </relatedItem>
        <relatedItem type="constituent">
            <titleInfo>
                <title>Bon Secours Market, Bon Secours Church, founded in 1658</title>
            </titleInfo>
            <location>
                <url>http://collections.ic.gc.ca/sva/qc/mon_01/images/img_09.jpg</url>
            </location>
        </relatedItem>
        <relatedItem type="constituent">
            <titleInfo>
```

Appendix C: Data Format XML Examples

```xml
            <title>Harbour, looking east</title>
        </titleInfo>
        <location>
            <url>http://collections.ic.gc.ca/sva/qc/mon_01/images/img_10.jpg</url>
        </location>
    </relatedItem>
    <relatedItem type="constituent">
        <titleInfo>
            <title>Ice shove of the St. Lawrence, Custom House</title>
        </titleInfo>
        <location>
            <url>http://collections.ic.gc.ca/sva/qc/mon_01/images/img_11.jpg</url>
        </location>
    </relatedItem>
    <relatedItem type="constituent">
        <titleInfo>
            <title>St. Peters Cathedral. Photolithograph</title>
        </titleInfo>
        <location>
            <url>http://collections.ic.gc.ca/sva/qc/mon_01/images/img_12.jpg</url>
        </location>
    </relatedItem>
    <relatedItem type="constituent">
        <titleInfo>
            <title>Inside Jesuits' Church</title>
        </titleInfo>
        <location>
            <url>http://collections.ic.gc.ca/sva/qc/mon_01/images/img_13.jpg</url>
        </location>
    </relatedItem>
    <relatedItem type="constituent">
        <titleInfo>
            <title>Christ's Church (Episcopal Cathedral)</title>
        </titleInfo>
        <location>
            <url>http://collections.ic.gc.ca/sva/qc/mon_01/images/img_14.jpg</url>
        </location>
    </relatedItem>
    <relatedItem type="constituent">
        <titleInfo>
            <title>Grey Nunnery</title>
        </titleInfo>
        <location>
            <url>http://collections.ic.gc.ca/sva/qc/mon_01/images/img_15.jpg</url>
        </location>
    </relatedItem>
    <relatedItem type="constituent">
        <titleInfo>
            <title>Presbyterian College, McGill College</title>
        </titleInfo>
        <location>
            <url>http://collections.ic.gc.ca/sva/qc/mon_01/images/img_16.jpg</url>
        </location>
    </relatedItem>
    <relatedItem type="constituent">
        <titleInfo>
            <title>Windsor Hotel</title>
        </titleInfo>
        <location>
```

Descriptive Metadata Guidelines for RLG Cultural Materials

```xml
            <url>http://collections.ic.gc.ca/sva/qc/mon_01/images/img_17.jpg</url>
        </location>
</relatedItem>
<relatedItem type="constituent">
    <titleInfo>
        <title>Lachine Rapids</title>
    </titleInfo>
    <location>
        <url>http://collections.ic.gc.ca/sva/qc/mon_01/images/img_18.jpg</url>
    </location>
</relatedItem>
<relatedItem type="constituent">
    <location>
        <url>http://collections.ic.gc.ca/sva/qc/mon_01/images/img_19.jpg</url>
    </location>
</relatedItem>
<relatedItem type="constituent">
    <location>
        <url>http://collections.ic.gc.ca/sva/qc/mon_01/images/img_20.jpg</url>
    </location>
</relatedItem>
<relatedItem type="constituent">
    <location>
        <url>http://collections.ic.gc.ca/sva/qc/mon_01/images/img_21.jpg</url>
    </location>
</relatedItem>
<relatedItem type="constituent">
    <location>
        <url>http://collections.ic.gc.ca/sva/qc/mon_01/images/img_22.jpg</url>
    </location>
</relatedItem>
<relatedItem type="constituent">
    <location>
        <url>http://collections.ic.gc.ca/sva/qc/mon_01/images/img_23.jpg</url>
    </location>
</relatedItem>
<relatedItem type="constituent">
    <location>
        <url>http://collections.ic.gc.ca/sva/qc/mon_01/images/img_24.jpg</url>
    </location>
</relatedItem>
<relatedItem type="constituent">
    <location>
        <url>http://collections.ic.gc.ca/sva/qc/mon_01/images/img_25.jpg</url>
    </location>
</relatedItem>
<relatedItem type="constituent">
    <location>
        <url>http://collections.ic.gc.ca/sva/qc/mon_01/images/img_26.jpg</url>
    </location>
</relatedItem>
<relatedItem type="constituent">
    <location>
        <url>http://collections.ic.gc.ca/sva/qc/mon_01/images/img_27.jpg</url>
    </location>
</relatedItem>
<relatedItem type="constituent">
    <location>
        <url>http://collections.ic.gc.ca/sva/qc/mon_01/images/img_28.jpg</url>
```

```xml
        </location>
    </relatedItem>
    <relatedItem type="constituent">
        <location>
            <url>http://collections.ic.gc.ca/sva/qc/mon_01/images/img_29.jpg</url>
        </location>
    </relatedItem>
    <relatedItem type="constituent">
        <location>
            <url>http://collections.ic.gc.ca/sva/qc/mon_01/images/img_30.jpg</url>
        </location>
    </relatedItem>
    <relatedItem type="constituent">
        <location>
            <url>http://collections.ic.gc.ca/sva/qc/mon_01/images/img_31.jpg</url>
        </location>
    </relatedItem>
    <relatedItem type="constituent">
        <location>
            <url>http://collections.ic.gc.ca/sva/qc/mon_01/images/img_32.jpg</url>
        </location>
    </relatedItem>
    <relatedItem type="constituent">
        <location>
            <url>http://collections.ic.gc.ca/sva/qc/mon_01/images/img_33.jpg</url>
        </location>
    </relatedItem>
    <relatedItem type="constituent">
        <location>
            <url>http://collections.ic.gc.ca/sva/qc/mon_01/images/img_34.jpg</url>
        </location>
    </relatedItem>
    <relatedItem type="constituent">
        <location>
            <url>http://collections.ic.gc.ca/sva/qc/mon_01/images/img_35.jpg</url>
        </location>
    </relatedItem>
    <identifier type="local">ONCA.b11310753</identifier>
</mods>
```

Descriptive Metadata Guidelines for RLG Cultural Materials

APPENDIX D: CHARACTER ENCODING

XML

Specification: Extensible Markup Language (XML) 1.1, www.w3.org/TR/xml11/

Legal characters for XML are the legal characters of Unicode, as well as tab, carriage return, and line feed.

A small number of characters are used as markup delimiters in XML (& [ampersand], > [greater than], < [lesser than], ' [apostrophe], and " [double quotation marks]), and must be represented by using numeric character references or specific strings when they appear in data content.

The UTF-8 and UTF-16 encodings of Unicode are both valid for XML. If your data is not in UTF-8 or UTF-16, XML requires that you identify your character set.

MARC Formats (MARC 21 and UNIMARC)

Specifications:
MARC 21 Specifications for Record Structure, Character Sets, and Exchange Media. www.loc.gov/marc/specifications/spechome.html

UNIMARC Manual: Bibliographic Format 1994 (including Update 3 2000). www.ifla.org/VI/3/p1996-1/sec-uni.htm

MARC 21 and UNIMARC are both implementations of ISO 2709. Both specify two mutually exclusive encodings: individual character sets with escape sequences (i.e., the code extension techniques of ISO/IEC 2022), or use of Unicode.

The individual character sets specified for use in MARC 21 are sometimes referred to as "MARC-8" (a reference to the 8-bit character sets shown in the MARC 21 Specifications). The MARC-8 character sets cover Latin, Arabic, Cyrillic, Greek, and Hebrew scripts, plus the East Asian Character Code (EACC) for Chinese, Japanese and Korean (CJK). The default character sets for MARC 21 records in a Latin script environment are ASCII and ANSEL (ANSI/NISO Z39.47). Non-Latin character sets used in a record are identified in Field 066 Alternate graphic representation.

The individual character sets specified for UNIMARC cover Latin, Cyrillic, Greek, African letters, Georgian, and Hebrew. The default character sets for UNIMARC are ISO/IEC 646 (i.e., ASCII) and ISO 5426. Other character sets registered with ISO may also be used. The character sets used in a record are identified in Field 100, subfield a, positions 26-29 and 30/33.

Unicode is specified for use in both MARC 21 and UNIMARC records. MARC 21 specifies use of the UTF-8 encoding of Unicode, indicated by the value "a" in Leader position 9. UNIMARC makes no statement about which Unicode encoding form or forms may be used in records, nor how to indicate the encoding form used.

Appendix D: Character Encoding

The MARC 21 Specifications include mappings to Unicode equivalents for the characters of the individual MARC-8 sets. Except for character sets that are common to MARC 21 and UNIMARC (ASCII, Cyrillic, and Greek), there are no officially published mappings to Unicode equivalents for the individual character sets of UNIMARC.

MARC 21 records encoded in UTF-8 should only include Unicode characters that have equivalents in the individual MARC-8 character sets. Technical specifications for the use of Unicode characters that do not have MARC-8 equivalents are being developed.

For more information see Ardie Bausenbach's article, "Character Sets and Character Encoding: A Brief Introduction," in *RLG DigiNews*, volume 8, number 2, www.rlg.org/en/page.php?Page_ID=17068#article2.