Z39.50 Profile for Access to Digital Collections

Draft Seven
(Final Draft for Review)

May 3, 1996
Library of Congress

Contents

(contents continued)

1. Background and Introduction

In August, 1995, the Library of Congress convened a team of representatives from several institutions to develop a Z39.50 profile access to digital libraries.

Early in its development the profile was renamed from Z39.50 Profile for the Digital Library Application to Z39.50 Profile for Access to Digital Collections (nicknamed the *Collections* profile) as its scope was narrowed, to apply to *navigation* of digital collections. Other groups were initiating independent efforts to develop profiles aimed at specific types of objects and collections. The intention was to coordinate these efforts and that these latter profiles would be developed as compatible extensions or subsets of the Collections profile. These latter profiles are referred to as *companion profiles*. An example is the CIMI profile: Z39.50 Application Profile Specification for Use in Project CHIO. CIMI, the Consortium for Interchange of Museum Information, is the sponsor of a demonstration project, CHIO (Cultural Heritage Information Online). Another example is the Z39.50 Profile for Access to Digital Libraries being developed at the Library of Congress.

1.1 Document History

The initial draft of this document was developed in early September, 1995; *Draft Two* was issued for discussion at the first meeting of the profile team, September 20. *Draft Three* was issued November 3, for the second meeting, November 14-15. *Draft Four*, issued in late December was prepared for a meeting that was subsequently canceled. *Draft Five* was developed via electronic mail and other discussion, for the meeting February 5-6, and was also distributed to the ZIG for discussion at the ZIG meeting February 7-9. *Draft Six*, issued March 18, reflected extensive discussion at the February 5-6 meeting, the February 7-9 ZIG meeting, as well as subsequent comments.

This draft, *Draft Seven*, reflects comments on Draft Six, and is available for review and comment until June 1. Following comment resolution, the profile will be proposed for approval.

1.2 Purpose of This Profile

This profile specifies a conforming subset of Z39.50-1995 to address problems described in section 1.3, in particular for access to digital collections organized via descriptive information whose structures are described within this profile. It provides semantics for navigating digital collections, to locate and retrieve objects of interest.

1.3 Problems Addressed by this Profile

Libraries and other institutions are creating a growing number of collections, organized thematically, for example, by subject, creator, historical period, etc.; with numerous, diverse objects, both digital and physical. These collections are often organized hierarchically and distributed across servers.

Significant resources may be invested in digitization and in the intellectual efforts of aggregation, organization, and description of the information in a collection. Yet to a remote user or client, the collection may appear to be simply an accumulation of objects and undifferentiated data, because there is no agreed-upon semantics for navigating the collection, to locate and retrieve objects of interest. Coherent organizational structures, imposed on the data, are necessary to provide a view that supports navigation.

A key obstacle to effective navigation is the inability to distinguish content from description. A primary goal of navigation is to locate and retrieve objects of interest; a vital step in that process is to locate relevant descriptive information. Thus it is useful to navigate among descriptive information as well as content, and consequently, to be able to distinguish content from description.

Various forms of descriptive aids have been developed for purposes of describing collections and objects, for example, finding aids, encoded archival descriptions, and exhibition catalogs. Often they do not have a well-defined structure and cannot be used alone for reliable navigation. At many institutions though, descriptive aids of these types have been created, at significant expense. It is therefore imperative that an application relying on descriptive information for collections and objects be able to utilize these available aids, rather than mandate the creation of new redundant structures. In order to exploit these descriptive aids, a higher level enveloping structure is necessary, to allows users and clients to navigate among and select descriptive aids of interest.

Another navigational problem is posed by the various and often complex relationships among and between objects and collections. For a given object, there may be other objects that are duplicates or variations (e.g. different resolutions) or which bear other relationships (e.g. surrogates). For any given collection there may be superior, subordinate, related, and context collections. These relationships among objects and collections must be modelled coherently.

Different digital objects, even within the same collection, may be retrievable by different protocols (e.g. Z39.50, HTTP, FTP). It is therefore important that structures describing these objects, and how they may be accessed, be defined independent of any specific protocol.

These problems amplify when a collection is distributed across institutions (or otherwise distributed in a manner requiring access to multiple servers). In particular some normalization of query semantics is necessary for coherent navigation of collections. Clients must be able to formulate queries that will be interpreted consistently across servers.

1.4 Scope Limitations

This profile assumes that *companion profiles* (compatible extensions to or subsets of this profile) will be independently developed, extending or limiting the use of this profile to specific applications or classes of information, for example, museum objects, satellite photographs, geospatial data, or chemical compounds. Thus this profile does not directly address all of the problems cited in 1.3, but provides a framework for companion profiles to do so:

- Digital Objects are treated as atomic, that is, their content is opaque and is not addressed by this profile. Thus the profile addresses searching descriptive information rather than searching Digital Objects. Companion profiles may model the content of specific objects (e.g., museum objects).
- Associated Descriptions (e.g., finding aids, exhibition catalogs; see 2.3.2) are treated as opaque (their content is not addressed by this profile), though clients may have at their disposal helper applications that are able to process or display them. Companion profiles may model the content of Associated Descriptions.
- The profile does not model complex relationships among objects of all classes. Companion profiles may do so for specific object classes.

Although this profile emphasizes the logical separation of descriptive information from content, it does not include guidelines or specifications for determining whether information is description or content. The profile does not address organizing principles, i.e, the means by which objects are aggregated into collections.

Although this profile addresses distributed collections, it does not address distributed databases. Different parts of a collection may be managed by different institutions; therefore different databases, corresponding to a single collection, may reside on different servers, but no individual database is distributed across servers.

Only the Z39.50 protocol is addressed by the profile. The profile does not, however, preclude multi-protocol clients, gateways, or digital collections where part of a collection is accessible via Z39.50 while another part is accessible by other protocols.

1.5 Definitions

Associated Description: A unit of descriptive information associated with a collection or object, for example, an encoded archival description or a finding aid (for an archival collections), or a catalog record (for library materials).

Child: Collection or object A is a child of collection B if B is a parent of A. A collection may have objects as well as subcollections as children.

Collection: A group of related objects and/or collections, possibly distributed across locations. A collection is a tree, where leaf nodes are objects and non-leaf nodes are subcollections.

Collection Descriptive Record: A Descriptive Record that describes a Collection.

Companion profile: An independently developed profile which is a compatible extension to or subset of this profile, extending or limiting the use of this profile to specific applications or classes of information, for example, museum objects.

Context Collection: Collection A is a context collection for collection or object B if it is a superior, related collection, and the organization with responsibility for the management of B considers that although collection A may be relevant to a user who is interested in B, any collection superior to collection A is likely not relevant.

Descriptive Record: A unit of descriptive information at a higher level of abstraction than an Associated Description. A Descriptive Record may include one or more Associated Descriptions in addition to other information that describes either a collection (and possibly its contents) or an object within a collection. A Descriptive Record is either a Collection Descriptive Record or an Object Descriptive Record.

Member (of a collection): Child.

Object: A Physical Object or a Digital Object.

Object Descriptive Record: A Descriptive Record that describes an object.

Parent Collection: Collection A is a parent of collection or object B if A is immediately superior to B.

Related Collection: Collection A related to collection or object B if the organization with responsibility for the management of B considers that collection A may be relevant to a user who is interested B.

Subcollection: Child collection.

Subordinate: Collection or object A is subordinate to collection B if A is a node on a tree whose root is B.

Superior Collection: Collection A is superior to collection or object B if B is a node on a tree whose root is A.

1.6 References

ANSI/NISO Z39.50-1995. *Information Retrieval (Z39.50): Application Service Definition and Protocol Specification*. See http://lcweb.loc.gov/z3950/agency.

CIMI Profile. *Z39.50 Application Profile Specifications for Use in Project CHIO.*

Z39.50 Profile for Access to Digital Libraries. Library of Congress.

2. Model, Assumptions, and Principles

The data model and descriptive data described in this section and section 4 are intended for general use, not specific to Z39.50. In particular, the intellectual content of the descriptive data modelled by this profile is distinguished from the format used for its transfer. Z39.50 specifics are described in section 5.

2.1 Model of an Object

An *object* is a *Physical Object* or a *Digital Object*. For any given Physical Object, there may or may not be a Digital Object which is its digital representation. Any given Digital Object may or may not be the digital representation of some Physical Object. A Digital Object may be text or it may be a digitized object; in any case, this profile treats a Digital Object as opaque.

2.2 Model of a Collection

A *collection* is a group of related objects and/or collections, possibly distributed across locations. Thus a collection is a tree, where leaf nodes are objects and non-leaf nodes are subcollections.

Collection A is said to be *related* to collection or object B if the organization with responsibility for the management of B considers that collection A may be relevant to a user who is interested B.

Collection A is said to be *superior* to collection or object B if B is a node on a tree whose root is A; B is said to be *subordinate* to collection A if A is superior to B.

Collection A is said to be a *parent* of collection or object B if A is immediately superior to B; B is said to be a *child* of collection A if A is a parent of B.

Collection A is said to be a *context* collection for collection or object B if it is a superior, related collection, and the organization with responsibility for the management of B considers that although collection A may be relevant to a user who is interested in B, any collection superior to collection A is likely not relevant.

The concept of a context collection is related to, but differs from, the concept of a root collection (which this profile does not employ). A context collection might informally be considered a "relative root" collection, for purposes of navigation. "Root" (as in "root collection") has the connotation of absolute or static. In contrast, "context" (as in "context collection") is relative; furthermore, although a collection might be a root collection from the view of a subordinate collection, a parent collection may subsequently be created for the former, so that it is no longer a root collection, and this new relationships might not be conveyed (nor might it be meaningful) to the subordinate collection.

For any given collection or object there may be zero or more *parent* collections, zero or more *related* collections and zero or more *context* collections. For any given collection there may be zero or more child collections.

A collection is a different logical construct than a database, in several respects:

- A database is an aggregation of records, and a collection is an aggregation of objects, some of which may be physical objects.
- The apportionment of records into databases might be based on very different criteria than the apportionment of objects into collections. For example, a server might provide the following databases: Books, Serials, Maps, Photos, SoundRecordings, and MotionPictures, containing digitized books, serials, maps, photos, sound recordings and motion pictures, respectively. A single collection might contain objects from all of these categories, and therefore records (corresponding to objects in the collection) would be distributed across these databases. The collection might be organized by theme, while the databases are organized, for example, by media type.
- The property of belonging to a collection differs from the property of belonging to a database; section 2.6 describes this difference.
- A database resides on a single server, while a collection may span servers (that is, different databases corresponding to a single collection may reside on different servers).
- A single database may include records corresponding to objects in more than one collection.
- A single object might belong to more than one collection, while the database record for that object may be retrievable from only one database.
- An object might belong to a single collection, while the database record for that object is retrievable from multiple databases, which may be on different servers.

2.3 Model of a Record

Z39.50 models database records, abstract database records, and retrieval records. It employs the concept of an abstract record structure defined by a database schema: a common understanding shared by client and server of the information contained in the records of a database, allowing logical components of that information to be addressed, selected, and represented.

A database record is a unit of information in a database, represented in a data structure local to the server. An abstract database record is an abstract representation of the information in a database record, formed by the application of an abstract record structure to the database record, where the abstract record structure is known to the client. A retrieval record is the information in an abstract database record represented in an exportable structure (by the application of a record syntax).

This profile employs the concept of a record according to the Z39.50 model, thus the term *record* is used to mean database record, abstract database record, or retrieval record, depending on usage and context. The Descriptive Record (Collection Descriptive Record and Object Descriptive Record) is a fundamental construct in this profile. The term *Descriptive Record* may refer to a database record, in particular when the context is searching. It refers to an abstract database record when the context is the definition of the Descriptive Record. It refers to a retrieval record when the context is retrieval.

2.4 Descriptive Information

This profile exploits organizational structures in order to allow a client to navigate through structured information, possibly hierarchically structured. A coherently defined set of descriptive data is used to manage and navigate collections of otherwise undifferentiated data. These organizational structures allow the data to be viewed as distributed collections, loosely hierarchical. (From a top-down view, a collections is a tree. From a bottom-up view, collections are directed graphs, because any object or collection may have multiple parents.)

2.4.1 Descriptive Information and Content

Collections modelled by this profile are characterized by the logical separation of descriptive information from content. This logical separation is not intended to constrain an implementation. Servers may choose to separate descriptive data from content in retrieval records using the record structure defined in this profile. However, servers are not required to perform this separation. In any case the profile does not prescribe nor address how servers store data.

In general, a collection may include Digital Objects that are completely self-describing, Digital Objects that include some self-description while additional descriptive information is separate, as well as Digital Objects where descriptive information is completely separate.

This profile does not provide guidelines for determining whether information is description or content. It is the responsibility of the institution managing a collection to make that determination, and this profile does not presume that different institutions will make the same determination for similar or even identical information. For example, a Physical Object, for which there is a digitized picture, could be modelled as a Physical Object with no corresponding Digital Object, where the picture is treated as descriptive information; alternatively, the picture could be treated as a Digital Object.

As another example, a Digital Object might be largely self-describing, including, for example, a caption that is an integral part of the Digital Object. A copy of that Digital Object might reside in a different collection, managed by a different institution, that might designate it instead to be description (of a physical object) with no corresponding Digital Object.

2.4.2 Modelling Descriptive Information

Collections and objects often have descriptive information associated with them. For example, archival collections are described in encoded archival descriptions and finding aids; library and other materials are described in catalog records; museum objects and collections are described in wall texts and exhibition catalogs.

An important design objective of this profile is to support the use of such descriptive information, here called an *Associated Description*, for use in navigating collections. A server implementing this profile may have created and stored an existing base of Associated Descriptions; it is the intention of this profile that the server be able to deliver these existing Associated Descriptions without the need to modify or reformat them.

Because of the wide variety of possible descriptive information, this profile does not attempt to enumerate or restrict the types of Associated Description that may be available. It is anticipated that companion profiles will enumerate the types of Associated Description appropriate for particular domains.

Neither does this profile attempt to describe the structure or contents of any Associated Description; for the purpose of this profile an Associated Description is an opaque data element passed from the server to the client along with other information relevant to navigating collections. A client that receives an Associated Description may either display it to the user or pass it to some application that can process it. The nature of such processing is outside the scope of this profile.

This profile supports the retrieval of information needed to navigate within and between collections and to retrieve both Digital Objects and Associated Descriptions. It does so by defining the structure of a Descriptive Record, which is at a higher level of abstraction than either a Digital Object or an Associated Description. A Descriptive Record may include one or more Associated Descriptions in addition to other information that describes either a collection (and possibly its contents) or an object within a collection.

This profile requires that a server be able to provide Descriptive Records (for collections and/or objects). These could be stored as database records, or alternatively, retrieval records may be created upon request by appropriate processing of other information available to the server, including information within digital objects and Associated Descriptions. This is, however, strictly an implementation matter; this profile does not specify how, or if, the Descriptive Records are to be stored. It only specifies the format for transfer from server to client. (Thus, although according to the Z39.50 abstract record model, a retrieval record corresponds to a database record, whether that database record physically exists on the server is transparent to the client.)

2.4.3 Collection and Object Descriptive Records

A Descriptive Record describes a collection or an object and is, respectively a *Collection Descriptive Record* or *Object Descriptive Record*.

A Collection Descriptive Record may provide an overall description of a collection as well as collective or individual descriptions of some or all of the objects in the collection.

An Object Descriptive Record describes an object: a Digital Object (which may or may not be the digital representation of a Physical Object) or a Physical Object.

Both a Collection Descriptive Record and Object Descriptive Record may list parent-, context-, and related- collections. A Collection Descriptive Record may list child collections as well.

For a given Digital Object, there need not necessarily be an Object Descriptive Record. The institution managing a collection to which the object belongs is responsible for making that determination. An object may, for example, be sufficiently described by a Collection Descriptive Record. As another example, a set of digitized photographs might be distinguished as distinct Digital Objects and aggregated into a collection or might be combined as a single Digital Object. If they are aggregated into a collection, there may be a single Collection Descriptive Record and there may or may not be individual Object Descriptive Records for some or all of the individual Digital Objects.

2.5 Datastore Model

A collection is represented by a *datastore*: part or all of one or more databases on one or more servers (distributed databases, however, are not addressed by this profile; see 1.4). To a given collection there corresponds a set of databases that include the descriptive records for that collection: Object Descriptive Records for objects and Collection Descriptive Records for subcollections.

The databases corresponding to a collection are not exclusive to that collection. They may include other records: records pertaining to other collections, or records not related to digital collections.

This model does not prescribe where Digital Objects reside; any given Digital Object may be represented as nested within its Object Descriptive Record, as a distinct record within the same database as its Object Descriptive Record, on a different database within the same server, or on a different server. A Digital Object may be accessible via Z39.50 or via another protocol.

2.6 Enumerating the Members of a Collection

For modelling purposes, *members* of a collection are objects and subcollections.

> Note: thus the property of belonging to a collection differs from the property of belonging to a database. For example, suppose an Object Descriptive Record describes a Physical Object in a collection. The Object Descriptive Record belongs to a database (but is not a member of the collection, because, by definition, only objects and subcollections are members) while the Physical Object is a member of the collection (but does not belong to a database).

A Collection Descriptive Record may provide an enumeration of the members of the collection it describes. Each item in the enumerated list is a pointer that refers to one of the following:

- an Object Descriptive Record (if the member is an object),
- a Digital Object (if the member is a Digital Object), or
- a Collection Descriptive Record (if the member is a subcollection).

3. Navigational Scenarios

This section describes some of the scenarios that sections 4 and 5 are intended to support.

When a client connects to a server providing access to digital collections, on behalf of a user interested in digital collections, the following scenarios are possible.

- The user might know the name of a collection of interest, as well as the database where the Collection Descriptive Record for the collection resides.
- The user may know the name of a collection but not the database where its Collection Descriptive Record resides. In that case, if the client and server both support Explain then the client may attempt to determine that database via Explain (see 5.5.)
- It may be that neither the client nor the user knows any collection names. In that case if the client and server both support Explain, the client might attempt to learn which databases in general correspond to collections and search those databases for desired Collection Descriptive Records. The client may then retrieve Collection Descriptive Records from these databases and display summary information to the user, including brief descriptions.
- The user may select a collection and then navigate to other collections of interest: the client may retrieve a list of related collections, including parent, superior, and context collections, brief descriptions of these collections, and descriptions of their relationship to the subject collection. The user might select one of these collections, determine its parent, superior, and context collections, etc.

In any case, eventually the user may identify a collection of interest. A number of possible scenarios might ensue:

- The client might retrieve an enumeration of the Associated Descriptions for the collection. For each Associated Description, the client may retrieve the Associated Description type, format, size, and a brief text description. The client might use the Associated Description types and format to determine which Associated Descriptions it can process or if not process, display to the user. The client might then display the brief descriptions (and size) of those Associated Descriptions to the user, who then selects an Associated Description of interest for the client to process or display.
- The client might retrieve from the Collection Descriptive Record an enumeration of the objects in the collection. For each object, the client retrieves a brief description for display to the user (if the number of enumerated objects is large, the client can retrieve a few at a time). For a given object:
    - The client might retrieve part or all of the Object Descriptive Record for the object, in particular an enumeration of its Associated Descriptions, and, as in the Collection Descriptive Record scenario above, the client displays brief descriptions of Associated Descriptions to the user, who selects an Associated Description of interest for the client to process or display.
    - Alternatively, the client might retrieve and display object metadata including a list of available formats for the object, and for each, its size, and level of integrity (see 5.4.9). The user might select a format and the client may then retrieve the object in that format.
- The client may search the collection based on user supplied search criteria. (Here, "search the collection" means search the databases corresponding to the collection, where the search is restricted to Object Descriptive Records which list the collection as a parent.) User supplied search criteria are not addressed by this profile, but may be addressed by companion profiles. If the client and server both support Explain the client may retrieve information about the databases, for example, attributes supported, in order to assist the user in formulating meaningful searches. (Note that this profile does not require support for Explain.)