Z39.50 Profile for Access to Digital Library Objects

Fourth Draft
August 15, 1996

The Z39.50 Profile for Access to Digital Library Objects (hereafter referred to as the *DL Profile*) is a companion profile to the Z39.50 Profile for Access to Digital Collections referred to as the *Collections Profile*. (Another companion profile to the Collections profile is the *CIMI Profile*: Z39.50 Profile for Museum Collections Information, developed by the Consortium for the Computer Interchange of Museum Information, for project CHIO: Cultural Heritage Information Online.)

1. Overview

As a Z39.50 Profile, the DL profile specifies a subset of Z39.50 features to support functional and user requirements for search and retrieval of information in digital library collections, specifically the Library of Congress digital library collections and similar collections.

The use of this profile might be one of several mechanisms used to access library digital objects. This particular mechanism is distinguished by the definition of an enveloping structure, called an Object Descriptive Record, which may (logically) encapsulate a digital object along with information describing the object. When the Descriptive Record does not (logically) encapsulate the object it describes, it instead provides a pointer to the object. In either case, in this profile, Z39.50 access to a digital object (for search or retrieval) is via its Object Descriptive Record.

1.1 Extensions to Collections Profile

The Collections profile includes several scope limitations, and delegates responsibility to companion profiles, to extend the scope in these areas. As a companion profile to the Collections profile the DL profile extends the collections profile in the following areas:

- The Collections profile treats digital objects as atomic and delegates to companion profiles any further modeling of a digital object. The DL profile provides a general model for a digital object; see 1.1.1. The profile also defines a schema corresponding to this model; see section 3.
- The Collections profile delegates to companion profiles the designation of categories of digital objects that the companion profile is intended to support. The DL profile defines categories of digital objects. See 1.1.2.
- The Collections profile treats Associated Descriptions as atomic and delegates to companion profiles the designation of categories of Associated Descriptions that the companion profile is intended to support. The DL profile defines categories of Associated Descriptions. See 1.1.3.
- The DL profile supports authentication, rights and permissions, access control, and resource control. See 1.1.4.
- The DL profile supports special characters within a search term, and language designation of the search term. See 1.1.5.

1.1.1 Model of a Digital Object

This profile provides a general and flexible model for the structure of a digital object. In this model, a digital object may consist of constituent parts, any of which may in turn consist of constituent parts, and so on. Constituent parts are represented as Z39.50 elements.

Consider a single digital object consisting of several images (e.g. photos or text images). Although the set of images comprises a single digital object, each must be distinctly representable and the object must convey the fact that there are distinct images, how many, and their individual characteristics. Thus they are represented as separate elements of a Z39.50 record.

Next suppose that the digital object not only includes a number of images, but also additional constituent parts, further structured; for example, each such constituent part may consist of several images. This introduces an intermediate level of aggregation.

The model of a digital object adopted by this profile assumes arbitrary levels of aggregation and is represented as a tree, where each non-leaf node has an arbitrary number of subtrees and/or leaves, and leaf nodes represent data.

Every node, whether a leaf or non-leaf node, has a string tag, whose purpose is to convey to the user what that node represents. A description (via a *description* meta-element) may also accompany any node, in case the string tag is not sufficiently descriptive.

This model could represent, for example, a digital object consisting of 10 boxes, each with 20 folders, each with 30 photos. String tags such as 'box', 'folder', and 'photo' could be used to convey the type of element (the type would be conveyed to the user, not the client; this profile does not attempt to define machine-processible content types). As a more complex example, a folder might include a variety of photos, maps, correspondences, etc. and perhaps the correspondences consist of several sequential digitized pages.

As another example, a digital object may consist of multiple volumes, each with a table of contents, several chapters, each chapter with sections, etc. String tags such as 'volume', tableOfContents', 'chapter', 'section', etc. could be used.

Repeating elements, and designation of the ordinal occurrence of an element among a set of repeating elements, is supported by this profile. For example consider an object with multiple "volumes", i.e. "volume 1", "volume 2" etc. "Volume 2" would be represented by the second occurrence of the element whose tag is 'volume'. "Third chapter", "fifth image" or "first five pages" would be similarly represented. Element specification eSpec-1 and record syntax GRS-1 provide these capabilities.

1.1.2 Categories of Digital Object

The profile defines the following categories of digital objects:

- Language-based
- Image-based
- Sound-based
- Motion-based

1.1.3 Categories of Associated Descriptions

The profile defines the following categories of Associated Descriptions:

- Cataloging Record (e.g. MARC record, Dublin core)
- Archival Register
- Header (file Header)
- Web Page

1.1.4 Authentication, Rights and Permissions, Access Control, and Resource Control

This profile supports:

- Authentication (at Initialization): capability for a client to submit a user id and password during Z39.50 initialization. See 5.4.
- Rights and Permissions metadata: capability for the server to return "terms and conditions" or "rights and permissions" information corresponding to digital information, as metadata along with the specific digital information. See 3.1.

- Access control: capability for the server to demand authentication (by prompting the client for authentication information) prior to processing a request for information. See 5.8.
- Resource Control: The capability for the server to notify the client of potential or actual resource requirements pertaining to a request for information, and prompt the client for permission to continue to process the request. See 5.9.

1.1.5 Character Set and Language Support for Search terms

This profile supports special characters within a search term, and language designation of the search term. For special characters, support is required for character set negotiation as specified in the Z39.50 Implementors Group (ZIG) Implementors Agreements. See http://lcweb.loc.gov/z3950/agency/agree.html. *[Details to be developed.]* For designation of the language of a search term, see 4.3.3. This profile supports search terms that read right-to-left (e.g. Arabic, Hebrew); see 4.1.1.

1.2 Pilot Collections

This profile is intended to provide access to the digital collections described briefly below. This list is intended as illustrative and is by no means exhaustive; it is used as a representative set of collections on which the specifications of this profile are based.

1.2.1 Detroit Publishing Company Collection

More than 25000 negatives, 20000 prints, 2900 transparencies, from the Detroit Publishing Company, 1880-1920. U.S. scenes (mostly), including buildings, towns, cities, universities, battleships yachts, resorts, natural landmarks, and industry.

Images available in four versions: GIF, TIFF thumbnail, reference JPG, uncompressed TIFF.

1.2.2 Nation's Forum Collection

Collection of 59 sound recording of speeches, made to preserve the voices of prominent Americans; 1918 and 1920. World War I topics, postwar issues, and the 1920 presidential election.

Also available for each speech:

- Text (ascii)
- Photo of the record label of the sound recording
- Photo of speaker

1.2.3 WPA Life Histories Collection

Life History Manuscripts from the Folklore Project, WPA Federal Writers' Project, 1936-40. 2900 documents from 300 writers from 24 states. 2000-15000 words each. 23000 page images total. Histories describe informant's family education, income, occupation, political views, religion and mores, medical needs, diet and miscellaneous observations.

Each document available in HTML, SGML (using American Memory DTD and Panorama viewer) and scanned page image (bitonal TIFF G3 or G4).

1.2.4 Finding Aid for Shirley Jackson Papers

Shirley Jackson was a master American short-story writer and novelist of the mid-20th century, best known for modern Gothic horror, in particular for her classic story, The Lottery, 1948. She also wrote stories about contemporary domestic life. Her papers, given to LC in 1967, consist of diaries, journals, correspondence, literary manuscripts, and miscellaneous papers. There are 7400 items, none digitized. The Finding aid is SGML tagged using the Encoded Archival Description standard, beta version.

1.2.5 Coolidge-Consumerism Collection

17,000 pages (images) of 1920s primary-source materials: manuscript, monograph, and serials Also photos and motion pictures. Documents various aspects of economic life in the U.S. during the 1920s. Includes the Calvin Coolidge Papers, focusing on the life of Calvin Coolidge during the six years he was president (1923-1929).

Scanned page images (bitonal TIFF G4) available for the contents of 152 manuscript folders selected from 14 manuscript collections. 73 folders have page images only; 79 have HTML and SGML also.

Photographs (170) available in GIF, TIFF thumbnail, reference JPG, and uncompressed TIFF. Text is available for 78 monographs and 56 serials.

1.2.6 Legislative Information System

Currently named THOMAS and available on the Web, the Legislative Information System (LIS) under development is a constantly growing collection of large, heterogeneous databases of legislative and legal information which includes the fulltext of the Congressional Record, Bills and Laws; Committee Reports and other documents; and Congressional Research Service products (e.g., Bill Digests).

Besides fulltext ASCII (searchable by boolean and relevancy-ranked queries) the expanded LIS will support SGML-tagged documents, PDF, and audio and video format standards for various data sets.

1.3 Metadata and Variants

A variant specification, metadata element (e.g. from tagSet-M or tagSet-G), or GRS-1 metadata field, may apply at any node (leaf or non-leaf) of the digital object tree. For any given such metadata component type -- a variant specification of a given class and type, metadata element of a given tag, or specific type of GRS-1 metadata -- the rules of applicability and inheritance are as follows; for any leaf node:

- If any such component is attached, it applies to that leaf (and that leaf only).
- If no metadata component of a given type is attached, then the most immediately superior occurrence, if any, of a metadata component of the same type applies; that is, among the superior subtrees with a component of that type attached, the one attached to the most subordinate subtree applies.

1.4 Representations of a Digital Object

A single digital object may have several *representations*, for example a "thumbnail", "highly compressed", "high resolution", "original", or "reference image". When these characterizations apply to the digital object as a whole, they are represented as (Z39.50) variants applied at the root of the object tree (i.e. at element 'root' of datatype Object; see 3.1.1). Representations may also apply at nodes subordinate to the root, and the rules of applicability and inheritance stated in 1.3 apply.

> *[Note: there is currently a proposal to add a new feature to variant-1, necessary to support this.]*

1.5 Z39.50 Access to Digital Objects

As in the Collections profile, a digital object may be accessed via Z39.50 or via some other protocol. For the DL profile however, when a digital object is accessible via Z39.50, it may be accessed via its Object Descriptive Record only. The use of Z39.50 to search or retrieve a digital object directly (not via its Object Descriptive Record) is not supported by this profile.

2. Descriptive Record Schema

The Collections profile specifies an abstract record structure for the Descriptive Record. The DL profile adapts that definition. The resulting abstract record structure, compatible with the Collections profile definition, is described below.

The main structure is as defined in 4.1 of the Collections profile, and is not repeated in the description below. The subsections of this section correspond to the subsections of 4.2 of the Collections profile; each describes a datatype derived from the main structure. When a datatype definition for *this* profile is identical to the corresponding definition in the Collections profile, the definition is not repeated (instead a reference to the Collections profile is given). When a definition is provided, the "repeatable" column might be omitted when that column is identical to the "repeatable" column in the corresponding definition in the Collections profile.

2.1 CollectionInfo

CollectionInfo is as defined in section 4.2.1 of the Collections profile.

2.2 ObjectInfo

Element	Occurrence/Value	Repeatable
typeOfObject	mandatory; value always = 1	no
categoryOfObject	mandatory; values: • ’dl:language’ • ’dl:image’ • ’dl:sound’ • ’dl:motion’ • ’dl:unspecified’	no
digitalObject	mandatory	no

2.3 AssociatedDescription

Element	Occurrence/Value
briefDescriptionOfAD	optional
categoryOfAD	mandatory; values: • ’dl:catalogingRecord’ • ’dl:archivalRegister’ • ’dl:header’ • ’dl:webPage’ • ’dl:unspecified’
description	mandatory

2.4 RelatedCollection

RelatedCollections is as defined in section 4.2.4 of the Collections profile.

2.5 CollectionAndDb

CollectionAndDb is as defined in section 4.2.5 of the Collections profile.

2.6 ServerAndDb

ServerAndDb is as defined in section 4.2.6 of the Collections profile.

2.7 EnumeratedMember

EnumeratedMember is as defined in section 4.2.7 of the Collections profile.

2.8 DigitalObject

Element	Occurrence/Value
actualDO	occurs if and only if pointerToDO does not occur; this is a structured element, presented according to the schema defined in section 3.
pointerToDO	occurs if and only if actualDO does not occur

2.9 Description

Description is as defined in section 4.2.9 of the Collections profile.

2.10 RecordPointer

As in the Collections profile, RecordPointer takes one of the following two forms:

(a) a database and record identifier; or

(b) some alternative identifier.

In case (a) it is assumed that Z39.50 will be used to search and retrieve the record. In case (b), some other protocol may be used. RecordPointer in this profile is as defined in section 4.2.10 of the Collections profile, with the following restriction: whenever RecordPointer takes form (a) it points to an Object Descriptive Record of a Collections Descriptive Record; it never points to a digital object. See 1.5.

3. Digital Object Schema

The Digital Object Schema is defined by this profile (an oid will be assigned), and pertains to the element actualDO within DigitalObject of the Descriptive Record Schema (2.8).

3.1 Tag Types

For this schema a GRS-1 record will use the following tagTypes:

TagType	Usage
1	Elements from tagSet-M defined in Z39.50-1995, Appendix TAG, TAG.2.1.
2	Elements from tagSet-G defined in Z39.50-1995, Appendix TAG, TAG.2.2.
3	String tags, for element 'child' of data type Tree.
4	Elements defined by this profile. No tagSet is registered by this profile because there is no other (known) profile that uses these elements. If a future profile wishes to do so, the tagset (implicitly) defined by this profile will be explicitly defined and registered.

3.2 Abstract Record Structure

An object is represented by a structured data element where the top level is a single occurrence of the TagSet-M element 13:

Element	Tagtype/tag	Datatype	Occurrence/value	Repeatable
object	1/13	Object	mandatory	no

Note: the definition of tagSet-M element 13 indicates the following: ".... tagSet-M elements that occur subordinate to this element apply only to that nested record." The reasons for its use here is to encapsulate a schema identifier (tagSet-M element 1) that will pertain to, and only to, the subordinate structure.

3.2.1 Object

Element	Tagtype/tag	Datatype	Occurrence/value	Repeat able
schemaIdentifier	1/1	Object Identifier	mandatory; oid of this schema	no
root	4/1	Tree	mandatory	no

3.2.2 Tree

Element	Tagtype/tag	Datatype	Occurrence/value	Repeatable
metadata	4/2	Metadata	optional	no
child	3/<string>	Tree or OCTETSTRING	mandatory	yes

3.2.3 Metadata

Element	Tagtype /tag	Datatype	Occurrence/value	Repeat able
global Identifier	4/5	International String	Optional; any given element might physically belong to more than one object; although this is otherwise transparent to the client, the client may use this identifier to determine that it has already retrieved this element.	no
brief Description	2/17	International String	Optional; additional description of subtree or leaf if string tag is not sufficiently descriptive.	no
element Ordering	4/6	International String	Optional, applicable only to non-leaf elements. How child elements are ordered.	no
creation Date	1/15	Generalized Time	optional	no
dateLast Modified	1/16	Generalized Time	optional	no
dateLast Reviewed	1/17	Generalized Time	optional	no

Notes:
(1) Terms and Conditions may also be included as part of Metadata, represented as GRS-1 ElementMetaData field 'usageRight'.
(2) For rules of applicability and inheritance, see 1.3.

4. Attributes

This profile specifies support for the following attribute sets: bib-1, collections-1, dl-1, and legis-1.

4.1 Support for Bib-1 Attribute Set

4.1.1 Truncation

The bib-1 Truncation attributes 'right truncation' and 'left truncation' should be supported, and in order to support terms independent of direction (i.e. left-to-right vs. right-to-left):

- 'right truncation' is to be interpreted to mean "truncation from the end of the string".
- 'left truncation' is to be interpreted to mean "truncation from the beginning of the string".

[Remainder of 4.1 to be developed.]

4.2 Support for Collections-1 Attribute Set

Support is required for the Collections-1 attribute set as specified in the collections profile, with the following exceptions:

- For Use attribute 'record-type' all term values are to be supported except 4 (record is a digital object).
- For Use attribute object-type, the only term value to be supported is 1 (object is a digital object).

4.3 dl-1 Attribute Set

This profile defines the attribute set dl-1 *[oid to be assigned]*.
The following attribute types are defined: Use, Context, CategoryOfAD, and Language.

4.3.1 Use Attributes

Value	Use	Term(s)
1	CategoryOfDO	as listed in 2.2

4.3.2 Context Attributes

Value	Context
1	Brief description
2	Associated Description
3	Digital Object

4.3.3 Category Attributes

Value	Category
1	Cataloging Record
2	Archival Register
3	Header

This is used in conjunction with a Context attribute of value 2; intended to restrict searching to within Associated Descriptions, and only those of the specified type.

4.3.4 Language Attribute

The value of the Language attribute specifies the language of the term. The attribute value is a string (requires the string attribute feature of version 3) indicating the language, based on Z39.53-1994.

4.4 legis-1 Attribute Set

This profile defines the attribute set legis-1 [*oid to be assigned].*

The following attribute type is defined: Use.

4.4.1 Use Attributes

Value	Use	Term(s)
1	CongressNumber	Integer; e.g. 101, for '101st congress'
2	CongressionalSession	Integer; value 1 or 2
3	LegislativeStatus	
4	LegislativeMaterialType	'bill', 'congressional record', committee report', 'committee hearing', 'CRS product', 'GAO Report', 'GAO Testimony', 'CBO report'
5	LegislativeBillNumber	InternationalString; e.g. 'h.r. 1425', 'S. 896', 'h.j.res. 125', 'sconres. 24', 'h.res. 99'
6	BillSponsor	InternationalString; personal name
7	LawNumber	InternationalString

5. Specific Profile Details

5.1 Conformance Level

This profile requires support for Enhanced Conformance as defined in section 6 of the Collections profile.

5.2 Protocol version

Support for version 3 of Z39.50 as specified by Z39.50-1995 is required.

5.3 Services and Facilities

Support for the Init, Search, and Present services is mandatory for origin and target. Support for Access-control and Resource-control services is mandatory for an origin.

5.4 Init

An origin is required to implement the IDAuthentication parameter of the Init request, using the format provided in the commented description of idauthentication within the Z39.50 ASN.1. The client should implement 'userId' and 'password' of 'idPass', as well as 'anonymous'. It is recommended that a server treat the absence of the 'idAuthentication' parameter as an implicit 'anonymous'.

5.5 Search

Support for the bib-1, Collections-1, and dl-1 attribute sets is mandatory, and support for legis-1 is required to search legislative collections (see 4).

5.6 Retrieval

5.6.1 Present Parameters

Support for Present parameter CompSpec is mandatory, and its usage is as specified in section 5.4.1 of the Collections profile.

5.6.2 Element Specification

Support for element specification eSpec-1 is mandatory. Its usage is as specified in section 5.4.10 of the Collections profile. In addition, support for 'wildThing' (within TagPath) is required. Support for 'compositeElement', optional in the Collections profile, is not required by this profile.

5.6.3 Variants

Support for variant set variant-1 is mandatory. A variant specification may occur:

- As a variant request in an element specification using eSpec-1; see 5.6.3.1.
- As an appliedVariant in a GRS-1 record; see 5.6.3.2.
- As a supportedVariant in a GRS-1 record; see 5.6.3.3.

5.6.3.1 Variant Request

Support for variant requests (within eSpec-1) is mandatory. A variant request will use the variant-1 definition and consist of one or more of:

- variantId (class 1);
- body part type (class 2);
- language and/or character set (class 4);
- piece (class 5): 'start' or 'next' of 'what-fragment wanted';
- meta data requested (class 6): size (type 2), variant list (type 5);
- 'no data' (class 9, type 1).
- 'variant description' (class 9, type 4). *[proposed new type]*

5.6.3.2 Applied Variant

Support for applied variants is required. An Applied variant may include:

• bodyPartType (class 2)	Always supplied.
• formatting (class 3)	Optional.
• language (class 4)	Optional.
• character set (class 4)	Optional.
• metaData -- size (class 7, type 2)	Optional, when accompanying data. Must be supplied for 'no data'.
• variant description (class 7, type 6)	Optional. *[proposed new type]*

5.6.3.3 Supported Variant

Support for supported variants is required. A supported variant may include:

• variantId	Always supplied.
• bodyPartType	Always supplied.
• formatting	Optional.
• language	Optional (if applicable).
• character set	Optional (if applicable).
• metaData -- size	Always supplied.
• variant description (class 7, type 6)	Optional. *[proposed new type]*

5.6.4 Record Syntax

5upport for GRS-1 is mandatory. Usage of record syntax GRS-1 within this profile is defined as in 5.4.11 of the Collection profile. metaData and appliedVariant must be supported. Use of ElementData is as specified in the Collections profile, and in addition, within ElementMetaData:

• hits	? (Use in a fashion similar to CIMI?)
• supported variants	Support mandatory. See 5.6.3.3.
• usageRight	Support mandatory.

5.6.5 Schemas

Support for the Descriptive Record schema (section 3) and Digital Object schema (section 4) are mandatory.

5.6.6 Brief and Navigation Descriptive Records

Support for element set names 'b' and 'navigation' (see 4.6 of Collections profile) are mandatory.

5.6.7 Direct and Indirect Access to Digital Objects

According to the Collections profile (see 4.2.8, data type DigitalObject) access to a digital object is provided either directly, via element actualDO, or indirectly via a pointer, element pointerToDO. In some cases a server has no choice but to provide a digital object indirectly: when the digital object is accessible not via Z39.50 but rather via a different protocol, or when the digital object is on a different server. The latter case is out-of-scope as this profile does not address distributed collections; however, the case of a non-Z39.50 access is recognized.

Thus, for this profile, when a digital object is accessible via Z39.50, it should be provided directly, that is, via element actualDO.

5.8 Access Control

Mandatory for origin, optional for target. *[Access control format to be developed.]*

5.9 Resource Control

Mandatory for origin, optional for target. *[Resource control format to be developed.]*