

The Alexandria Digital Library Project: Metadata Development and Use

Linda L. Hill and Greg Janée
lhill@alexandria.ucsb.edu and gjanee@alexandria.ucsb.edu
Alexandria Digital Library Project
University of California, Santa Barbara

Introduction

The Alexandria Digital Library (ADL) Project has the unique advantage of developing georeferenced digital libraries holding both textual and geospatial resources and providing services that provide geospatial description and access for all resources. The ADL Project evolved from the Map & Imagery Lab (MIL) of the University of California, Santa Barbara, Davidson Library and is today an operational part of the MIL [1]. Computer science, geography, and library science disciplines have been involved in ADL's development which has been funded primarily by the National Science Foundation (NSF). ADL's metadata structures and knowledge organization systems (KOS) are grounded in practices originating in traditional libraries, in geospatial and data clearinghouses, and in geographic information system (GIS) analysis environments. "Geospatially-aware" digital library software, collections, and demonstration projects have emerged from this work that are freely available for wider use. This chapter presents a high-level view of ADL's metadata and KOS activities from the project's ten years of developing the infrastructure and components for *distributed* georeferenced digital libraries.

Georeferenced Digital Libraries

There are two fundamental types of georeferencing: by placenames and by geospatial coordinates (primarily longitude and latitude but also by grid reference such as the Universal Transverse Mercator, or UTM, system). Placenames are used in discourse and text, subject headings and index terms, labels on maps, and to identify administrative districts for addresses, statistics, and data. Geospatial coordinates are used to represent the location of features on the surface of the Earth and the coverage of maps, aerial photographs, remote-sensing images, and datasets of various kinds. Typically, the predominate use of placename or coordinate place referencing has been associated with domains of information resources that have been treated by separate information management systems: text-based systems on the one hand and GIS on the other.

Georeferenced digital libraries, by allowing users to pursue a place-based interest starting with either a placename or a map location and retrieving both textual documents and geospatial data, bridge the gap between these types of information resources. This is achieved by integrating digital gazetteers into the system to translate between placenames and coordinate locations and by accommodating coordinate-based spatial referencing in basic metadata structures, data storage, data processing, and services.

In traditional libraries, geospatial documents and information services have been the focus of map collections and map librarianship [2]. This is again a form of the separation in treatment that generally exists for textual documents versus geospatial documents and data. Coordinate fields for MARC cataloging were developed by map librarians and are currently available in the 034 (coded cartographic data) and 255 (cartographic

mathematical) fields [3]. As their names imply, these fields were intended for “cartographic” description and that is how they are used. It is not current practice to use these fields to catalog documents such as environmental impact reports that are also explicitly associated with coordinate-defined locations; placename referencing using MARC field 651 (geographic name) and to some extent field 650 (topical term) is used instead.

Geographic information systems (GIS) provide a rich modeling and analytical processing environment for geospatial data, with complex methods of geospatial referencing, data manipulation, and visualization. They also include placename referencing in the form of labels for identified features within a GIS layer, e.g., the name for a lake in a hydrographic layer. In GIS, the coordinates are the primary focus and the placenames are secondary, the exact opposite of the text-based systems. Metadata documentation of the geospatial aspects of geographic datasets is very detailed to support both evaluation of fitness for use for particular purposes (e.g., can the data be used for navigation?) and for computer processing (e.g., geodetic datum, scale of resolution, level of certainty, scheme for coordinate representation).

Georeferenced digital libraries merge the mathematical (coordinate representation) and the textual (placenames references) georeferencing into an integrated system, where crossover in representation allows a user, using a single query, to find and accumulate information of all types that are associated with a geographic location and to visualize the results on a map. In addition, ADL is designed to be a distributed system of stand-alone

or networked nodes. It is important to note that ADL's emphasis for metadata development has not been at the item level of description, but on metadata to support distributed searching and access across dissimilar collections, as well as on the design of KOS (i.e., gazetteers and thesauri) needed for georeferenced digital libraries. The ADL specification has no requirements for item-level description but rather accommodates whatever item-level metadata the collection owner chooses to use through the mapping of searchable data elements to search indexes, as explain later in this chapter.

Overview of ADL Metadata and KOS Developments

A full description of the ADL distributed library architecture is outside the scope of this chapter; see [4]. Briefly, ADL is structured as a set of distributed nodes (peers) each supporting one or more *collections of items*, subsets of which may be “published” (made visible) to other nodes. The combined *library* is the sum of all such collections.

Collections are documented by collection-level metadata. Items are largely undefined by the architecture beyond the expectation that they are independent items documented by item-level metadata and that they are uniquely identified within a particular collection.

While the contents of the distributed library may be heterogeneous, three features integrate the library into a uniform whole: (1) the ADL “bucket” system (a common model for search metadata); (2) a common format for collection-level metadata; and (3) a central collection discovery service that operates on indexes created from the collection-level metadata. Each ADL node is implemented as a Java web application that provides Java, HTTP, and RMI (Java remote method invocation) client interfaces. A second interface layer defines the interface between ADL nodes and collection implementations,

or “drivers.” ADL provides several drivers that support different collection implementation strategies, including relational databases, XML document stores, and Z39.50 protocol services (experimental as of November 2003).

The Alexandria Digital Library (ADL) Project website is the main access point for all ADL research and operational documentation and links to related resources [5]. To provide an overview for the discussion that follows, specific ADL metadata and KOS activities for distributed georeferenced digital libraries are listed here, with associated references.

Metadata:

1. Support for search across dissimilar distributed collections through a common set of attributes that constitute search indexes [6]
2. Design of collection-level descriptions which support discovery of appropriate collections, and provide contextual and inherent information about collections for end users [7]
3. Design of metadata for access; that is, standard views of item-level metadata describing what users/clients need to know to fetch and process it [8]
4. Development of a simple spatial geometry language, based on the Geometry Markup Language (GML) [9]
5. Collaborative development of the metadata design for the Digital Library for Earth System Education (DLESE) [10]
6. Development of a content standard for computational models [11; 12]

Knowledge organization systems (KOS):

7. Design of protocols for the searching of distributed gazetteers and thesauri [13; 14]
8. Design of a content standard for gazetteers [15]
9. Development of a Feature Type Thesaurus (FTT) for gazetteers [16]
10. Use of concept spaces in support of undergraduate classroom education [17; 18]

Metadata Structures to Support Distributed Searching and Retrieval

Key metadata components of the ADL architecture are the search buckets, collection-level metadata, standard reports, access metadata, and browse image metadata (Figure 1).

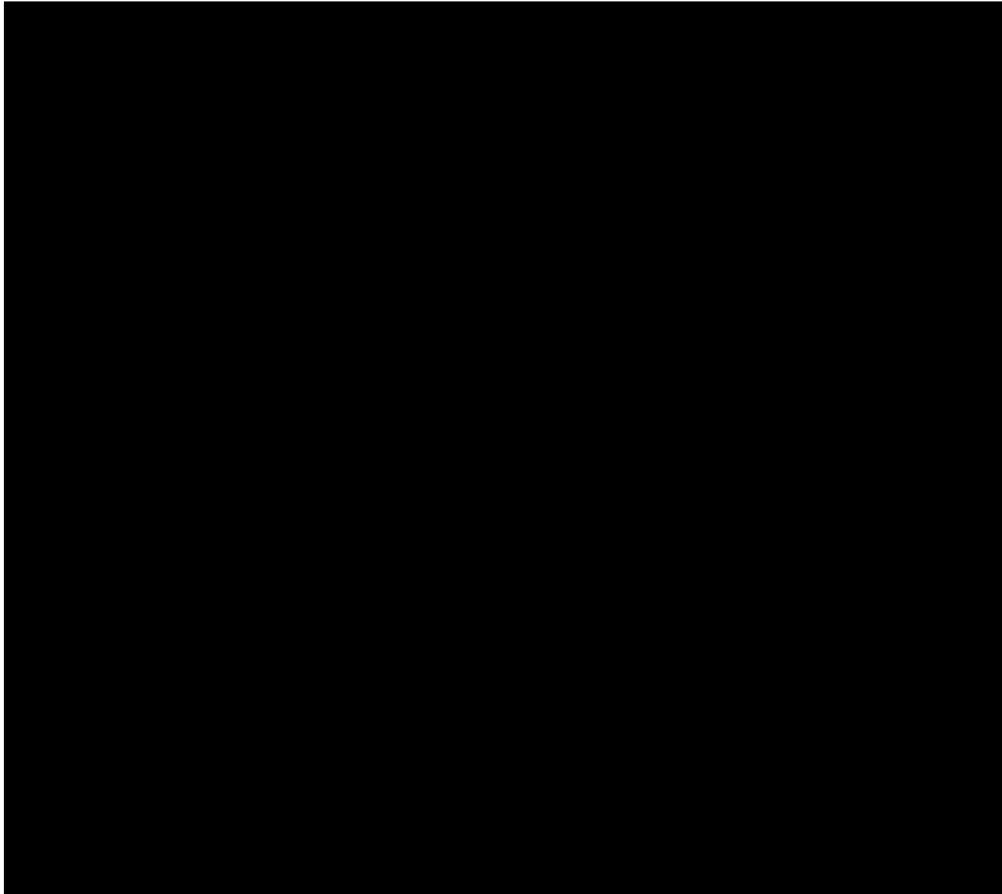


Figure 1. Metadata component of the ADL distributed library architecture

The *search bucket* component is specifically designed to support common search indexes for distributed, heterogeneous collections (Figure 2) [6].

Name	Definition	Data Type	Notes (DC = Dublin Core)
subject-related text	Text indicative of the subject of the item, not necessarily from controlled vocabularies.	textual	Title and assigned term buckets are sub-buckets to subject-related text. Mappings to this bucket can include abstracts, notes, free-text keywords, etc. Closest DC element: DC.Subject
title	The item's title(s).	textual	Closest DC element: DC.Title
assigned term	Subject-related terms from controlled vocabularies.	textual	Closest DC element: DC.Subject
originator	Names of entities related to the origination of the item.	textual	Both persons and organizations; no particular syntax required. Closest DC elements: DC.Creator & DC.Publisher
geographic location	The subset of the Earth's surface to which the item is related; expressed as a geometric region and defined in WGS84 latitude/longitude coordinates, expressed in an ADL-defined language.	spatial	Closest DC element: DC.Coverage.Spatial
coverage date	The calendar dates to which the item is relevant expressed according to ISO 8601.	temporal	Closest DC element: DC.Coverage.Temporal
object type	The intellectual type of the item.	hierarchical	Closest DC element: DC.Type.
feature type	The type of the feature for gazetteers specifically.	hierarchical	Closest DC element: DC.Type
format	The physical type of the item.	hierarchical	Closest DC element: DC.Format

identifier	Names and codes that function as unique identifiers with, optionally, associated namespaces.	identification	Closest DC element: DC.Identifier
-------------------	--	----------------	-----------------------------------

Figure 2. Core ADL Search Buckets

These search buckets are based on library practices and on common indexes to bibliographic databases and they can be compared to the Dublin Core Metadata Initiative (DCMI) in that they are composed of high-level attributes common to most item-level descriptions [19]. A key feature of the ADL *search bucket* metadata structure is that it is composed of *typed* data elements for which specific *search constraints* can be defined (Figure 3).

Bucket type	Value type	Constraints	Example
spatial	Any of several types of geometric regions defined in WGS84 latitude / longitude coordinates, expressed in an ADL-defined syntax.	overlaps contains within	Find maps with coverage within this search region 
temporal	Range of calendar dates or single date in ISO 8601 syntax.	overlaps contains within	Find aerial photos that have a date within the range of 1925 and 1935
hierarchical	Term from a cited scheme; that is, a controlled vocabulary or thesaurus.	<i>is a</i> relationship (hierarchical expand)	Find objects that are “images” and include objects that are <i>narrower terms</i> of “images” also.
textual	Text	any words (OR) all words (AND) phrase	Find placenames containing both “santa” and “barbara”.
numeric	Real number in standard scientific notation.	<, =, >, ...	Find populated places with population greater than 500,000.
identification	Identifier with an optional associated namespace.	matches	Find a book with ISBN 0-201-63274-8.

Figure 3. ADL Search Bucket Types

The types are spatial (coordinates), temporal, hierarchical term sets, textual, numeric, and identifiers. In addition to being typed, the buckets are semantically defined to ensure expected search results and to guide the mapping from the item-level metadata to the search buckets. There are no constraints or expectations for the metadata structure used to describe the items in the collections. Mappings from item-level metadata to the search buckets, which are done by the collection owner, are expected to meet the semantic definitions of the buckets and the mappings themselves are documented explicitly in the collection-level metadata.

Alexandria Digital Library's *collection-level metadata* provides a structure to describe the inherent and contextual attributes of a collection, here defined as any collection of items formally presented as a collection within a digital library [20]. Contextual information includes title, responsible party, scope and purpose, update frequency, etc. Inherent information is gathered from the collection itself, such as total number of items and subtotals by type, format, year and decade, and spatial coverage (which can be displayed on a map). Contextual information also includes the documentation of the mappings made from selected item-level metadata elements to the ADL search buckets. This mapping is very specific, for instance: MARC 21 fields [100](#) (main entry---personal name) and [110](#) (main entry---corporate name) have been mapped to the Contributor bucket. This collection-level documentation is useful in three ways. It provides:

- User understanding of the scope and origin of the collection;

- More specific search capability by querying item-level descriptive elements directly; that is, extending the search capability to a two-level process of searching first at the level of common, high-level search buckets and subsequently at the level of item metadata elements; and
- Collection discovery by focusing queries on the collections most likely to contain the information needed.

The third metadata component for the distributed digital library architecture, the *access metadata*, is for the purpose of documenting intellectual property rights and what a user needs to know to retrieve library objects. It formalizes the description of the aspects of the item needed for downloading and processing [21]. Included are the basic descriptive elements of title, description and rights, and the key attributes needed for downloading, accessing through a programmatic protocol or service, connecting to a web interface for more information, or obtaining an off-line item.

The fourth component is a metadata structure for *standard reports* so that the results obtained from heterogeneous collections are returned to the user or client in a predictable format. Standard reports reflect the reduced set of information about the item in the search buckets. Extended reports for the full item-level metadata are not specified by the ADL framework; the collection owner provides these according to the metadata structure used for the items or by local processes that create different metadata views of the item-level records.

A final metadata component required is one for describing the *browse images* that accompany items in the collections. These reduced images are included in the standard and extended reports to show the user a view of the object (Figure 4). The metadata for browse images includes an identifier and URL for the image and details about the format, display size in pixels, and transfer size in bytes.

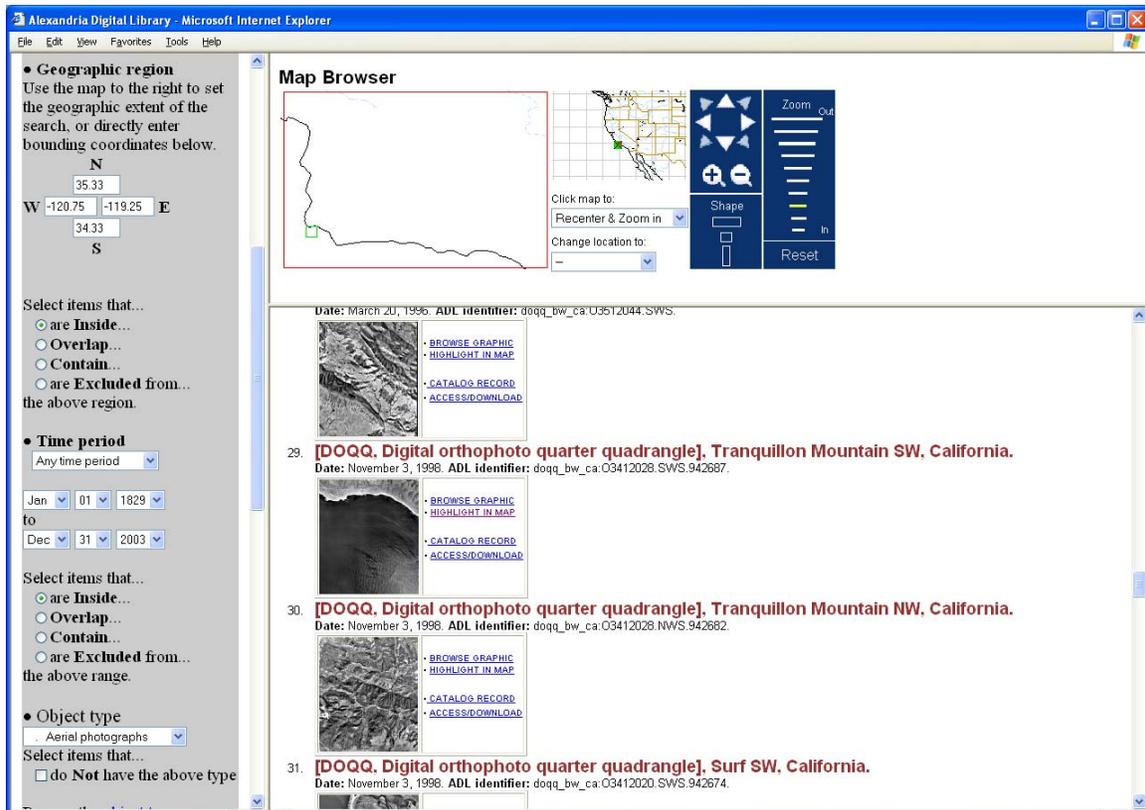


Figure 4. Screen shot of an ADL interface showing query parameters on the left, map browser on the top, and browse graphics with results for the search.

In summary, the ADL Project has designed a distributed library architecture which includes metadata structures to support searching on core descriptive elements across heterogeneous collections, and collection-level metadata that defines the collections for human understanding and provides statistics and specifications needed for processing and

collection discovery. These metadata components include geospatial representation as well as common core elements of item-level description, such as title, subject, and the persons, organizations, and dates associated with the item. Additional metadata structures are defined for accessing the items themselves, specifying standard reports, and describing the browse images that accompany image objects. All of these have been refined through the implementation of the operational Alexandria Digital Library.

Metadata Structures for Special Purposes

Metadata frameworks for representing geospatial data, at the item level, exist within and for the GIS community. The U.S. Federal Geographic Data Committee (FGDC) published version 1 of their *Content Standard for Digital Geospatial Metadata* (CSDGM) in 1994 and the second version in 1998 [22]. The Australian New Zealand Land Information Council (ANZLIC) published version 2 of their spatial information metadata guidelines in 2001. [23] The International Standards Organization Technical Committee 211 (ISO TC 211) has recently released the *Geographic Information Metadata* standard [24] that evolved from the FGDC and ANZLIC standards. Profiles of the ISO standard will be developed to harmonize the earlier geospatial metadata standards with the ISO standard. Another source of geospatial metadata specifications is the Open GIS Consortium (OGC), “an international industry consortium of 258 companies, government agencies and universities participating in a consensus process to develop publicly available geoprocessing specifications” [25]. The *Geography Markup Language* (GML) is “an XML grammar written in XML Schema for the modeling, transport, and storage of geographic information” [26].

These standards establish the framework for the item-level description of geospatial data. They are, however, developed for mapping and analysis, and for the reuse of geospatial data beyond the original purpose for the collection of the data. For use in digital library environments, something simpler is needed that preserves valid geospatial representation but simplifies the metadata structure. In collaboration with the Digital Library for Earth System Education (DLESE) and NASA, ADL developed a set of metadata elements to describe geospatial location as part of the ADN metadata framework [10]. This same set of geospatial elements was used for the *Content Standard for Computational Models* [12; 11]. For the DCMI, Simon Cox developed a set of metadata elements to describe geospatial points and bounding boxes, with the same goal of encouraging the inclusion of geospatial referencing in metadata outside of the GIS environment [27; 28].

Greg Janée, lead software engineer for ADL, developed the following principles to guide the development of a new “simple geography language” that will be based on existing standards and will be suitable for use in digital library metadata [9]:

1. The number of different shapes supported by the language must be large enough to support both description of object footprints with reasonable fidelity and spatial searching over those footprints, but should be as small as possible to lessen the burden on users of the language. Simple shapes under consideration include points to represent point features (e.g., water wells); boxes as a simplified way to represent area features; disks (areas described by point and radius) to represent areas with uncertain location and extent; simple polygons (i.e., no donut

- shapes) to represent the shape of areas with greater fidelity than is possible with boxes; and polylines to represent linear features such as rivers. Aggregations of sets of shapes, either constrained to one kind of shape (e.g., set of points, set of polygons) or unconstrained, may also be needed.
2. The spatial reference system (SRS) in which primary shapes are defined (i.e., the coordinate system + datum) must not be mandated by the language, but should be declarable in a standard way. Mandating the use of a particular SRS places too high a burden on the users of the language, who would be forced to translate from a SRS used locally to the SRS mandated by the simple geometry language. Such translations can be mathematically complex and introduce unintended consequences.
 3. The geometry language must provide a *lingua franca* that virtually all geometry consumers and producers can understand; in practice, due to simplicity of implementation and widespread support, this means that bounding boxes (also known as minimum bounding rectangles) must be encoded in parallel with all primary shapes. Such bounding boxes must be defined with respect to a spherical topology (i.e., there must be no discontinuity at the 180° meridian) and, notwithstanding principle 2, must be defined in a standard SRS, namely, WGS84 latitude/longitude coordinates. (SRS translations are generally not a problem when converting to simplified forms such as bounding boxes.)

In this statement, “datum” refers to system of measurement and representation that defines the size and shape of the earth and the origin and orientation of an Earth

coordinate system. There are many of these definitions designed for different purposes. The World Geodetic System of 1984, WGS84, is one such geodetic datum that is often used. The requirement that a bounding box be associated with each primary shape is needed to support base-level geospatial searching using *greater than* and *less than* comparisons of coordinate values in systems with limited geospatial query matching capability.

The OGC's *Simple Features – SQL* specification [29] and its XML companion specification, GML version 2 [30], provide much of what the above principles require -- with the exception of principle 3, mandating a simple *lingua franca*. As of November 2003, a preliminary version of a “simple geography language,” based on the GML specification, has been developed. Further development, implementation, and testing of the specification within ADL will continue before being published for general comment and use.

The metadata design challenge for non-GIS metadata is to incorporate a simplified representation of geospatial location that is consistent with the standards developed by the geospatial data community. There is an educational challenge also: to help collection managers develop an understanding of geospatial location as a key representational component and a key place-based retrieval parameter across multiple domains of knowledge. The ability to find and accumulate information about a location, including the geography, climate, occurrence of biological organisms, cultural history and artifacts,

remote sensing images, photographs (aerial and ground-based), social statistics, and descriptive documents, depends on meeting these challenges.

Representing Named Geographic Places

Toponymy is the “taxonomic study of place-names, based on etymological, historical, and geographical information” and the term is also used to mean the placenames for a region – e.g., the toponymy of El Salvador [31]. Toponymic authorities exist at national and state/province levels which apply established rules to designate the official names for geographic locations. In the United States, this authority is the U.S. Board on Geographic Names which is an “interagency board established by public law to standardize geographic name spellings for use in U.S. Government publications” [32]. At the United Nations, a permanent Commission known as the “United Nations Group of Experts on Geographical Names” promotes “consistent use worldwide of accurate place names” through cooperative efforts among countries [33].

Placenames (e.g., Chicago, the Rocky Mountains, Coit Tower) are the dominant means of georeferencing outside of the geospatial data environment. Only some placenames are authorized for official use by toponymic authorities. Other names are known as colloquial names (e.g., “the Windy City” and “the Rockies”), or variant names. Different toponymic authorities can make different names official for the same place and, of course, placenames and other aspects of named geographic places change through time.

Gazetteers are dictionaries or indexes of named geographic features (places), and geospatially-defined gazetteers contain entries where geographic features are defined by

both placenames and geospatial location. Such gazetteers can be used to translate from a placename (official or not) to a map location. For example, a document or image or a query statement about Chicago can be linked to the map location of Chicago through the use of a gazetteer. Once the coordinate location is known, a spatial search for the geographic area of Chicago can be extended to retrieve relevant information that does not contain “Chicago” as a searchable placename. Gazetteers, therefore, are key components of georeferenced digital libraries.

Each toponymic authority, atlas, GIS system, and gazetteer reference book uses its own *ad hoc* set of descriptive elements, data structures, and type categories for gazetteer data. The Alexandria Digital Library, therefore, developed a *Gazetteer Content Standard* (GCS) as a general descriptive standard for gazetteer data and used it to build a worldwide gazetteer of 4.4 million entries by combining the data from the two U.S. federal gazetteers, plus other datasets. A web-based client for the ADL Gazetteer, with a map browser for search and display, has been available since 1996 and is currently used to support a wide variety of applications worldwide [15; 34]. The GCS is structured as an XML schema. In the summer of 2003, a relational database model for version 3.1 of the GCS was developed and implemented in PostGreSQL, a freely available database management system with support for spatial data. In addition to the GCS, a specification for the *ADL Gazetteer Protocol* has been published and implemented to support searching across distributed gazetteers of various structures [13].

These descriptive structures -- the GCS and the protocol's standard report -- represent two ends of the spectrum from complex to simple (Figure 5).

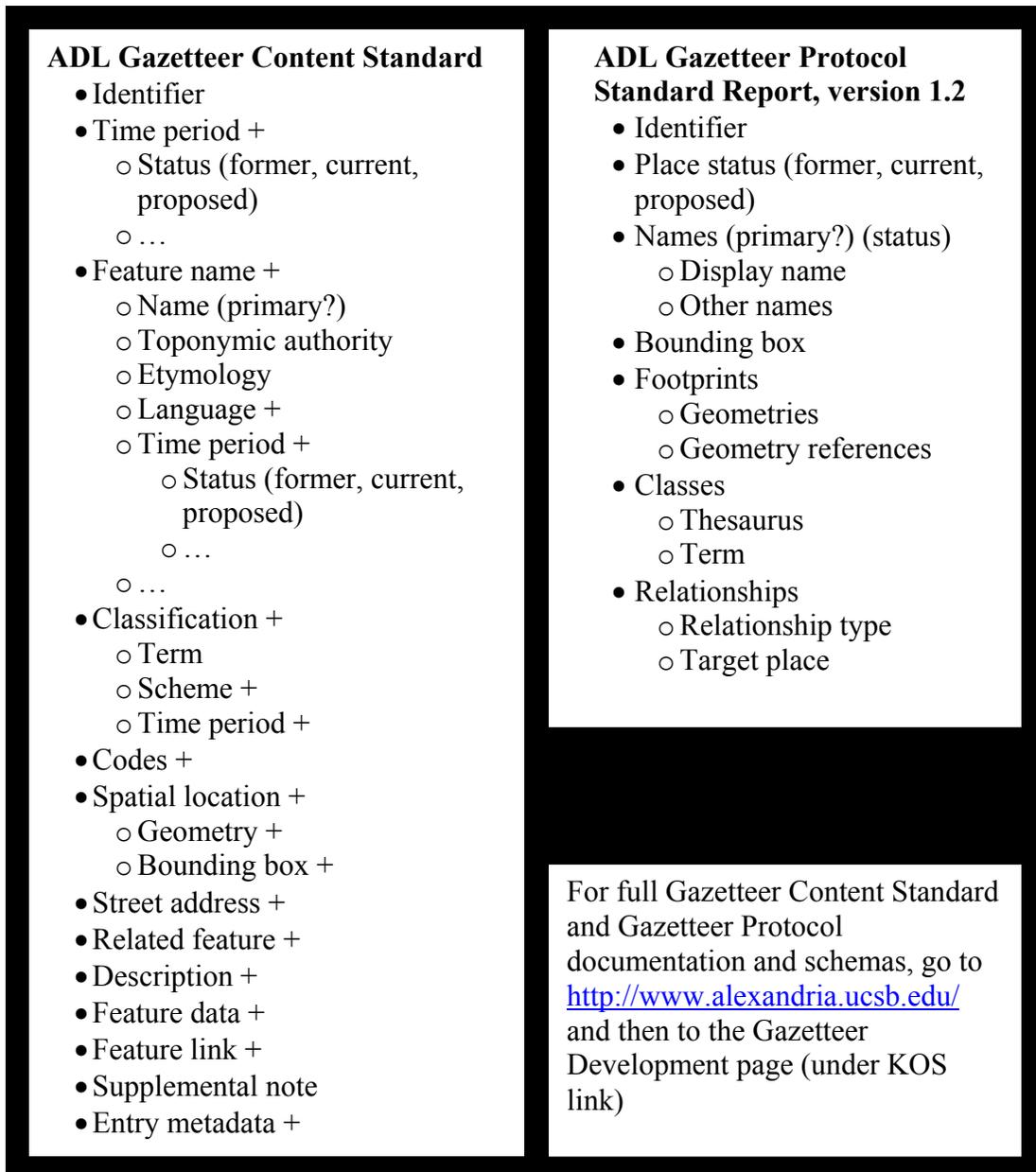


Figure 5. Top-level view of the *ADL Gazetteer Content Standard* with comparison to the report specification of the *ADL Gazetteer Protocol*, to illustrate the complexity of one and the simplicity of the other.

The GCS is structured to accommodate

- Complexity of placenames

- Authoritative vs. colloquial names
- Multilingual representations
- Variant spellings
- Explanation of the origin of the name (etymology)
- Uncertainties of our knowledge about ancient placenames;
- Varying geospatial representations of location (e.g., points, bounding boxes, and polygons and geospatial data from various sources);
- Primary and secondary type categories from specified typing schemes; and
- Complexity of time periods for name use, geospatial extent, and administrative relationships.

The GCS is also designed to attribute pieces of information about a place to a particular source of that information, so that a single entry can be composed of information gathered from multiple sources.

The *Gazetteer Protocol*, on the other hand, uses an abstract model of a gazetteer containing the key descriptive components needed to support distributed searching (similar to the ADL search bucket approach). The protocol can be used with any structure for the gazetteer data.

The ADL GCS is a *metadata-like* or catalog approach to representing named geographic places. That is, there are *records* representing each named geographic place and, within each record, explicit relationships can be declared with other gazetteer entries. For example, the record for Chicago, Illinois can include the fact that it is *part of* the state of

Illinois (another entry in the gazetteer). The spatial footprints for Chicago and Illinois also establish this fact spatially; the explicit statement establishes it administratively. The two U.S. federal gazetteers use a catalog record approach [35; 36], as does the ISO draft standard for gazetteers – “Spatial referencing by geographic identifiers” [37].

There are other familiar structures used for gazetteers. One is the hierarchical thesaurus structure, as used by the *Getty Thesaurus of Geographic Names* [38]. In a thesaurus structure, the administrative hierarchy is the organizing principle for the *whole-part* hierarchy. New places are inserted into the hierarchy by the editorial staff. Other structures are the simple list, or the authority list, which may not include the geospatial location, the table of *feature labels* for a GIS data layer linked to polygons or points, and the index by placename in the back of an atlas linked to map numbers and grid references.

There are several advantages to the catalog approach to gazetteers. One is that the hierarchical relationships can be more extensive than just the administrative hierarchy. A place can be said to be *part of* a watershed or a geological basin or an economic zone in addition to the administrative hierarchy. The catalog approach is more suited to *open contribution* models also, since new places do not have to be inserted into an existing hierarchy. For some purposes, however, the thesaurus approach based on administrative relationships is most suitable.

Other applications for the gazetteer model have become apparent since its development. It can also be applied to the description of other *named entities* having *spatial-temporal* definitions. One application is for the description of named time periods (e.g., the Iron Age or the Cretaceous Period) where the primary definition is a range of dates rather than geospatial coordinates. Another application is for the description of named spatial-temporal events, such as hurricanes or expeditions, where both the time range and the spatial coordinates are key components of the definitions.

Developing Thesauri and Concept Spaces

This book and this chapter are focused on metadata developments, but some mention should be made of the development of knowledge organization systems (KOS) that are used with the metadata structures to describe information objects. For ADL, this includes the development of the *Feature Type Thesaurus* (FTT) [16], the *ADL Thesaurus Protocol* [14], and the development and use of concept spaces to represent the important scientific reasoning and knowledge concepts for undergraduate teaching [17].

The FTT formalizes a set of terms (e.g., “populated places”, “mountains”, “towers”) within a hierarchical structure and is used to categorize gazetteer entries. It was developed because no shared set of such categories exist – only the *ad hoc* set of classes used by individual gazetteer creators [16]. The FTT terms are used as *values* for the *class* element in the GCS, and the FTT has been adopted or adapted for other purposes as well.

The *ADL Thesaurus Protocol* is designed to support the searching of multiple, distributed thesauri. It defines an abstract model of thesauri and the search operations that can be performed on them. In summer 2003, this protocol has been implemented on a test basis only and is one of several such approaches being evaluated in a search for a general solution to the problem of integrating a variety of KOS resources into digital libraries [39].

The concept-based approach to teaching and learning has been prototyped at the University of California, Santa Barbara by developing an undergraduate course in Physical Geography and presenting it in two academic terms [18; 17]. This approach integrates a collection of illustrations and models with the domain-specific concept space and with lecture composer software. The end result is a multi-screen presentation that can be navigated with either the 2-dimensional concept space or the lecture outline during classroom presentations. The components are also available to students for out-of-class study. It is a potentially powerful way to transmit to students the important concepts and the structure of a field of knowledge. Evaluation of the effectiveness of this approach is continuing.

Lessons Learned

The ADL Project, through its two phases of NSF digital library funding and its implementation as an operational georeferenced digital library, has provided a rich testing environment for digital library development. Along with software for distributed digital libraries and collection building, the project has accumulated a wealth of experience in designing and implementing metadata structures and KOS resources for georeferenced

digital libraries. The research, development, and implementation process continues, but at this point in time some observations can be made about what succeeded and how the ADL approach differs from other metadata and digital library initiatives.

The ADL Project has demonstrated how geospatial description and access can be integrated into digital libraries to provide users access to all types of information and data about a geographic location – overcoming not only data format differences and operationally distinct information services, but also hindrances such as multi-lingual, multi-cultural, multi-temporal, and colloquial placename variances.

In the process of developing software and metadata support for a distributed digital library architecture, ADL demonstrated the need to create metadata structures specifically designed for searching across dissimilar collections. Formal specification of collection-level metadata permits discovery of collections likely to have information relevant to a query. In the ADL search buckets and the gazetteer and thesaurus protocols, data typing and associated search constraints support a high level of search capability across collections that can be very diverse.

ADL's creation of metadata structures for interoperability across dissimilar collections demonstrates that the benefits of interoperability can be achieved without specifying a particular metadata standard for item-level description. However, it is also clear that customization of item-level metadata for the requirements of a particular set of objects should be based, as much as possible, on the use of shared metadata elements sets,

successful metadata design practices, and shared KOS resources to make interoperability more successful. Increasingly, metadata structures will be composed of element sets that have been developed as components to cover certain aspects of description. Geospatial representation is one of those areas.

The ADL Project has used standards from formal standards bodies when stable standards exist and can be adopted. At the same time, ADL has formalized and published the metadata, KOS, and distributed library structures developed through the course of the project as an informal standards effort. A key to success for these efforts is that the resulting product has incorporated the best of what is already in place and is consistent with or uses the emerging frameworks for the web and knowledge organization. A measure of the value of the informal standards process is the extent to which the resulting frameworks are adopted or adapted by emerging communities of practice. Both formal and informal standards development is needed. In areas of emerging practice, the informal path moves quicker and is informed by actual prototyping and test conditions.

It is a simplistic expectation that everyone is capable of creating good metadata or KOS resources but it is an expectation that has been prevalent to the detriment of predictable search and retrieval across distributed collections. The drive to create metadata and term lists automatically is also misguided if it doesn't include a human-in-the-loop for quality assurance and feedback to the system so that it can improve. Metadata and KOS design and creation must be a team effort where the expertise of domain experts and computer engineers is integrated with the expertise of metadata and information retrieval experts.

This implies, among other adjustments, a shifting of educational requirements and employment opportunities for the profession of librarianship.

The ADL Project consists of multi-disciplinary teams meeting on the common ground of digital library development: computer scientists/engineers, geographers, and library/information specialists primarily. This is a challenging environment where knowledge, assumptions, expectations, and language use is often not shared but is implicitly expected. It is a creative environment, absolutely required for digital library development, but the difficulties should be acknowledged. One problem is that terms with shades of meaning or explicit specification in one domain are sometimes used generically outside of that domain. To relieve the frustration, some mutual learning needs to happen and, if necessary, troublesome terms need to be used only with definition.

Conclusions

Geospatial access to and analysis of data and information is a key discriminating factor in many fields – to list a few: biodiversity, national security, transportation, social science, politics, cultural history, urban planning, epidemiology, natural resource exploration, and emergency management. Metadata and information service design and implementation is at a critical juncture where these activities must integrate geospatial georeferencing with traditional practices so that users can locate place-based information from documents and data, from libraries and data centers, and from museums and web pages. Two paths to doing this are (1) to incorporate coordinate representation directly into metadata instances for all types of information objects and into distributed digital library services, and (2) to

develop distributed gazetteer services to translate placenames in subject headings and text documents into coordinate locations, either as an aid to cataloging (metadata creation) or as a component of information retrieval services. Implementing geospatial access services in cataloging and information retrieval services through the translation of placename references to coordinate references using gazetteer services is an obvious development path. With sufficient development, such services will ease the cost and level of effort of extending geospatial access to all types of information.

Acknowledgements

The ADL Project has been funded primarily by National Science Foundation as part of their digital library initiative: NSF IR94-11330 (phase I) and ISI-98-17432 (phase II).

ADL metadata development is a team effort involving members of the ADL

Implementation Team and the staff of the Map & Imagery Lab, Davidson Library,

University of California, Santa Barbara.

Notes

1. Alexandria Digital Library Project. (2003). Operational ADL Home Page. Retrieved September 9 2003, from the World Wide Web: <http://www.alexandria.ucsb.edu/adl/>.
2. Larsgaard, M. L. 1998. *Map librarianship : an introduction*. 3rd ed. Englewood, Colo.: Libraries Unlimited.
3. U.S. Library of Congress: Network Development and MARC Standards Office. (2003). MARC standards. Retrieved September 9 2003, from the World Wide Web: <http://lcweb.loc.gov/marc/>.
4. Janée, G., and Frew, J. 2002. The ADEPT Digital Library Architecture. Paper given at ACM-IEEE Joint Conference on Digital Libraries, July 14-18, 2002, at Portland, OR.
5. ADL Home Page. Alexandria Digital Library Project. Retrieved September 9 2003, from the World Wide Web: <http://www.alexandria.ucsb.edu>.

6. Janée, G., Frew, J., Hill, L. L., and Smith, T. R. 2001. The ADL Bucket Framework (extended abstract). Paper given at Third DELOS Workshop on Interoperability and Mediation in Heterogeneous Digital Libraries, September 8-9, 2001, at Darmstadt, Germany.
7. Alexandria Digital Library Project. (2003). ADL Collection Metadata. Retrieved September 9 2003, from the World Wide Web:
<http://www.alexandria.ucsb.edu/~gjaneecollection-metadata/>.
8. Janée, G., Frew, J., and Valentine, D. 2003. Content Access Characterization in Digital Libraries. Paper given at ACM-IEEE Joint Conference on Digital Libraries, May 27-31, at Houston, Texas.
9. Janée, G. 2003. Principles for a simple geometry language for use in digital library metadata. Internal document, September 13, 2003, Alexandria Digital Library Project, University of California, Santa Barbara.
10. Digital Library for Earth System Education. (2003). ADN Metadata Framework. Retrieved September 20 2003, from the World Wide Web:
<http://www.dlese.org/Metadata/adn-item/index.htm>.
11. Crosier, S. J., Goodchild, M. F., Hill, L. L., and Smith, T. R. 2003. Developing an infrastructure for sharing environmental models. *Environment and Planning B: Planning and Design* 30:487-501.
12. Hill, L. L., Crosier, S. J., Smith, T. R., and Goodchild, M. F. 2001. A content standard for computational models. *D-Lib Magazine* 7 (6).
13. Janée, G., and Hill, L. L. (2002). ADL Gazetteer Protocol. (Version 1.1). Alexandria Digital Library Project. Retrieved May 2 2003, from the World Wide Web:
<http://www.alexandria.ucsb.edu/gazetteer/protocol/>.
14. Janée, G., Ikeda, S., and Hill, L. L. (2003). ADL Thesaurus Protocol. (Version 1.0). Alexandria Digital Library Project. Retrieved September 9 2003, from the World Wide Web: <http://www.alexandria.ucsb.edu/thesaurus/protocol/>.
15. Hill, L. L. (2003). Guide to the ADL Gazetteer Content Standard, version 3.1. Retrieved September 20 2003, from the World Wide Web:
<http://www.alexandria.ucsb.edu/gazetteer/ContentStandard/version3-1/GCS3-1-guide.doc>.
16. ———. (2002). Feature Type Thesaurus. Alexandria Digital Library Project. Retrieved September 6 2003, from the World Wide Web:
http://www.alexandria.ucsb.edu/gazetteer/FeatureTypes/FTT_metadata.htm.
17. Smith, T. R., Ancona, D., Buchel, O., Freeston, M., Heller, W., Nottrott, R., Tierney, T., and Ushakov, A. 2003. The ADEPT concept-based digital learning environment. Paper given at Research and Advanced Technology for Digital Libraries, 7th European Conference, ECDL 2003, at Trondheim, Norway.
18. Smith, T. R., Zeng, M. L., and ADEPT Knowledge Team. 2002. Structured models of scientific concepts for organizing, accessing, and using learning materials. Paper given at Seventh International ISKO Conference. 10-13 July 2002, Granada, Spain.
19. Dublin Core Metadata Initiative. (2003). Dublin Core Metadata Initiative (DCMI) Home Page. Retrieved September 11 2003, from the World Wide Web:
<http://www.dublincore.org/>.

20. Hill, L. L., Dolin, R., Frew, J., Janée, G., and Larsgaard, M. 1999. Collection Metadata Solutions for Digital Library Applications. *Journal of the American Society for Information Science (JASIS)*. *Special Topic Issue on Metadata* 50 (13):1169-1181.
21. Alexandria Digital Library Project. (2003). ADL Access Report DTD. Retrieved September 9 2003, from the World Wide Web: <http://www.alexandria.ucsb.edu/middleware/dtds/ADL-access-report.dtd>.
22. U.S. Federal Geographic Data Committee. (1998). Content Standard for Digital Geospatial Metadata. (Version 2). Retrieved June 1 2001, from the World Wide Web: <http://fgdc.er.usgs.gov/metadata/contstan.html>.
23. Australia New Zealand Land Information Council - Spatial Information Council. (2001). ANZLIC Metadata Guidelines. (version 2). Retrieved September 20 2003, from the World Wide Web: <http://www.anzlic.org.au/download.html?oid=2358011755>.
24. International Organization for Standardization (ISO) - Technical Committee 211. 2003. Geographic Information - Metadata.
25. Open GIS Consortium Inc. (2003). OGC Home Page. Retrieved September 20 2003, from the World Wide Web: <http://www.opengis.org/>.
26. ———. (2003). Geography Markup Language (GML) Implementation Specification. (version 3). Retrieved September 9 2003, from the World Wide Web: <http://www.opengis.org/techno/documents/02-023r4.pdf>.
27. Cox, S. (2000). DCMI Box Encoding Scheme: specification of the spatial limits of a place, and methods for encoding this in a text string. Dublin Core Metadata Initiative. Retrieved April 26 2003, from the World Wide Web: <http://dublincore.org/documents/2000/07/28/dcmi-box/>.
28. ———. (2000). DCMI Point Encoding Scheme: a point location in space, and methods for encoding this in a text string. Dublin Core Metadata Initiative. Retrieved April 26 2003, from the World Wide Web: <http://dublincore.org/documents/dcmi-point/>.
29. Open GIS Consortium Inc. (1999). OpenGIS Simple Features Specification for SQL. Retrieved November 13 2003, from the World Wide Web: <http://www.opengis.org/docs/99-049.pdf>.
30. ———. (2002). OpenGIS Geography Markup Language (GML) Implementation Specification. (version 2.1.2). Retrieved November 13 2003, from the World Wide Web: <http://www.opengis.org/docs/02-069.pdf>.
31. Encyclopedia Britannica : entry for "toponymy" from online public website. Retrieved November 3 2003, from the World Wide Web: <http://www.britannica.com/eb/article?eu=74822>.
32. U.S. Board on Geographic Names. (1998). Retrieved November 3 2003, from the World Wide Web: <http://geonames.usgs.gov/bgn.html> and <http://earth-info.nima.mil/gns/html/bgn.html>.
33. United Nations Group of Experts on Geographic Names. Retrieved November 3 2003, from the World Wide Web: <http://unstats.un.org/unsd/geoinfo/englishUNGEGN.pdf>.

34. Alexandria Digital Library Project. (2003). ADL Gazetteer Client. Retrieved September 20 2003, from the World Wide Web:
www.alexandria.ucsb.edu/clients/gazetteer.
35. U.S. National Imagery and Mapping Agency. (2003). Geonet Names Server. Retrieved September 20 2003, from the World Wide Web:
<http://164.214.2.59/gns/html/>.
36. U.S. Geological Survey. (2003). Geographic Names Information Service (GNIS). Retrieved September 20 2003, from the World Wide Web:
<http://mapping.usgs.gov/www/gnis/>.
37. International Organization for Standardization (ISO). 2002. Geographic information: Spatial referencing by geographic identifiers.
38. Getty Information Institute. (2003). Getty Thesaurus of Geographic Names On Line. Retrieved September 20 2003, from the World Wide Web:
<http://www.getty.edu/research/tools/vocabulary/tgn/>.
39. Hill, L. L., Buchel, O., Janée, G., and Zeng, M. L. 2002. Integration of Knowledge Organization Systems into Digital Library Architectures: Position Paper. Paper given at Thirteenth ASIS&T SIG/CR Workshop on "Reconceptualizing Classification Research", November 17, at Philadelphia, PA.