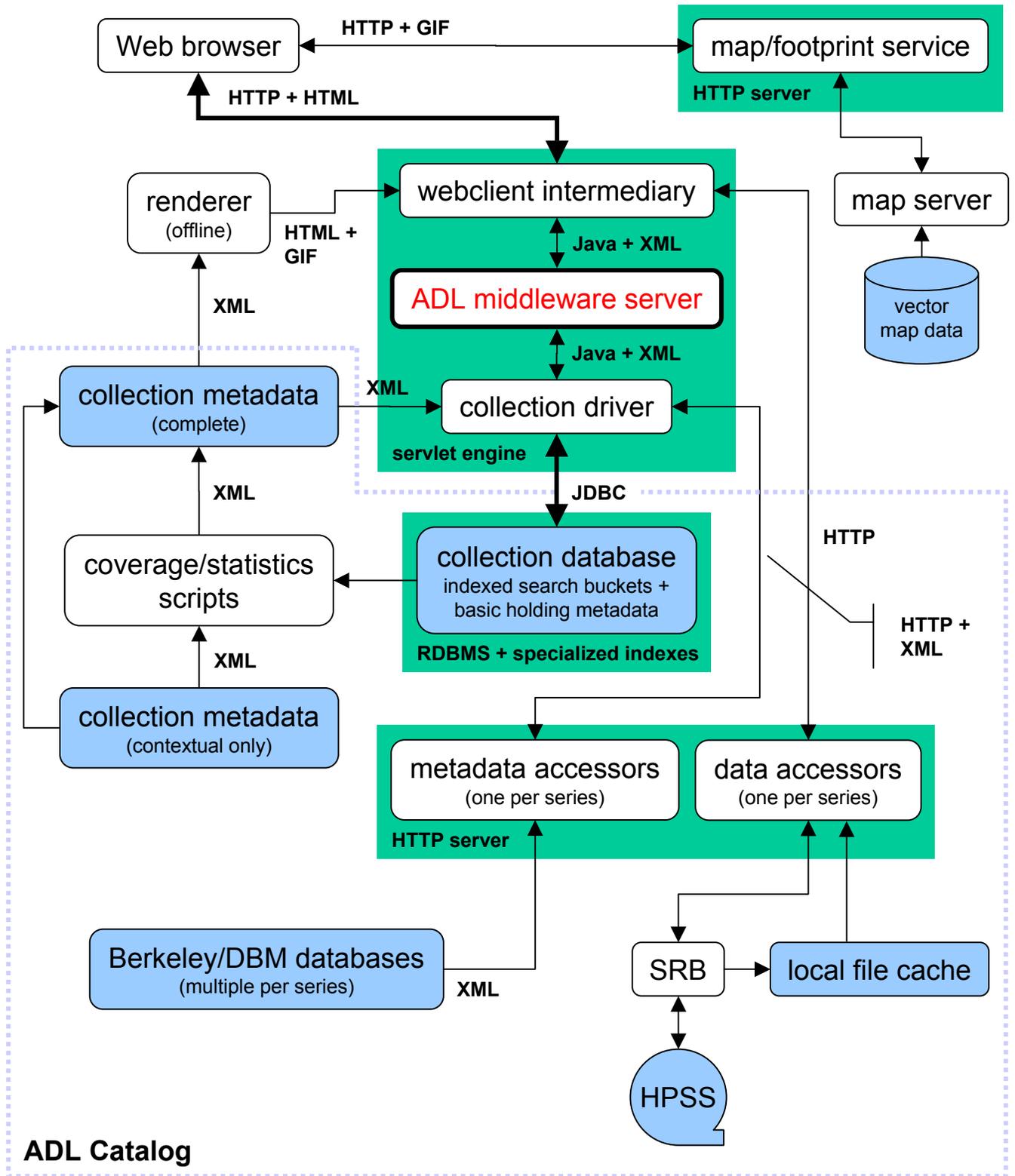


Alexandria Digital Library

and the ADL Catalog, a representative collection



The Alexandria Digital Library (ADL) is a generalized library system developed with the specific purpose of giving clients simplified, uniform access to heterogeneous collections of geospatial data and other georeferenced materials. Current collections within ADL include the locally-developed ADL Catalog and ADL Gazetteer; other, remotely-managed collections are under development. The ADL Catalog is a union-type collection of geospatial data (digitized maps, aerial photographs, remote-sensing datasets, etc., both online and offline) and bibliographic references to non-digital georeferenced materials. It currently contains approximately 2.5 million items and requires 1.5 terabytes of storage. The ADL Gazetteer contains 4 million geographic place and feature names, with associated geographic locations and historical and technical metadata. Clients to ADL include a simple interface that runs within a standard Web browser and a proxy client that makes ADL available from the SDLIP information retrieval protocol. The Web browser interface to ADL is freely available to the UC system and can be accessed at <http://webclient.alexandria.ucsb.edu/>.

Figure X above/below shows the major components that make up the library and a representative collection within the library, the ADL Catalog. Due to space limitations, ADL itself has been abstracted down to a single box in the diagram (*ADL middleware server*) and only the Web browser interface to ADL is shown. Also, the lower half of the diagram—the collection-specific portion—is effectively replicated for each collection. Also not shown are the systems that manage collections and ingest new items from source materials.

The ADL middleware server is implemented as a set of Java objects supporting multiple invocation interfaces. In the context of the Web browser client, the objects are accessed as servlets, and thus ADL runs within a servlet engine. As a general rule, all inter-system communication in the library uses HTTP, and all ADL interfaces and ADL-related metadata use XML encoding according to ADL-defined DTDs. The user's Web browser communicates with ADL via a *webclient intermediary*, which converts XML to HTML and also makes up for a number of deficiencies characteristic of Web browsers. The user's Web browser also makes use of an ADL-provided service that generates background maps and footprint displays.

Both the data and the functional interfaces that comprise a collection such as the ADL Catalog are distributed across various, closely-related servers and systems. Central to the collection is a *collection database*, which holds both basic item-level metadata and the indexes that support searching. The database resides in an RDBMS with extensions to support the specialized features ADL requires such as geospatial searching. Closely related to the collection database are multiple *Berkeley/DBM databases* that store pre-computed item-level metadata reports; *collection-level metadata*, which is discussed below; and *item-level data files*, which are stored in native form in a tape archive (HPSS) at the San Diego Supercomputer Center, and which are locally cached in a disk farm at UCSB.

The collection is accessed through a variety of mechanisms. ADL accesses the collection database using JDBC. Item-level metadata reports are accessed through CGI scripts running under an HTTP server; the scripts in turn use custom software to access the Berkeley/DBM databases. Item-level data files are accessed via the webclient intermediary through additional CGI scripts; these scripts in turn use the Storage Resource Broker (SRB) to access HPSS. The latter scripts also provide data-level access control and manage the local file cache.

The collection-level metadata is generated from static, contextual metadata (scope and purpose, maintaining agency, contact, etc.) and dynamic, intrinsic metadata such as numbers and types of items in the collection, spatial distribution of the items, and so forth. The intrinsic metadata is computed automatically from the collection database by a set of data aggregation and visualization scripts. The collection-level metadata is made available in XML form and also rendered into human-readable HTML with inline GIF images.

Access control is supported at three places in the architecture: server-level, collection-level, and data-level. Which access control schemes are employed at these three places is configurable; currently, ADL employs IP- and password-based access control at the server level; no access control at the collection level; and, at the data level, IP-based access control on certain series of items as required.

Persistence of identifiers is accomplished in an ad hoc manner by using virtual, function-oriented names for all systems and services, and by using one-time accession numbers to refer to collection items.