Society of Archivists Data Standards Group

A Guide to Archival and Related Standards
Standards applicable to archives; for the digital delivery of repository guides, finding-aids, and images of material from collections.

Title

Name of Standards Developing Organisation
The Library of Congress (Z39.50 Maintenance Agency)

Current version

Replaces

Abstract
Z39.50 is a standard for specifying a client/server-based protocol for searching and retrieving information from remote databases, and is therefore of relevance to many working with information resources in a distributed environment.

Description
Z39.50 became an approved standard by the American National Standards Institute (ANSI) in 1988. It is a protocol designed to enable communication between different computer systems. It is typically used to enable a user to search across a number of database resources simultaneously, such as library and archives catalogues, and it can be applied to a whole range of database resources.

The actual process is based upon the client/server model: a client sends a request for information to a server and the server responds. The client may be called the ‘source’ or the ‘origin’ and the server is the ‘target’ of the request.

An example of a Z39.50 enabled system is the Arts and Humanities Data Service (AHDS), which can interrogate across all AHDS Service Providers' catalogues. Many databases, such as Medline, Copac and Zetoc, provide Z39.50 access; they may all be included in a federated search, maybe a search via a library portal or using a bibliographic product such as EndNote.
Using Z39.50, a user can enter a query at their own PC terminal and gather results from a whole range of resources. For the user this process appears seamless - they can effectively cross-search different systems, or targets, by one simple search. The essence of a successful search via Z39.50 is to successfully manage the 'state' of the connection between the systems.

Z39.50 includes a number of 'Services' which relate to the types of operation that can be carried out between the client and the server. The three most basic services are Initialization, Search and Present. The client/origin of the query makes itself known to the server/target of the query through the initialization process (init.) to establish effective communication. The search is then sent from the client and the target responds. The client can also provide more specific instructions about how the information should be presented, such as requesting the data in a certain format (e.g. MARC, Dublin Core, EAD) or limiting the number of records that are returned.

The values for the set-up of the two systems, which define the parameters of the transaction, are governed by attributes. Use attributes define the types of search (title, author, date, etc.). Other attributes define values for settings like less than, greater than, and field positions. Different systems may have different values set for these attributes, and may not necessarily support all Z39.50 Services, which can impair the success of a remote search. Defining attribute sets can help with this and a number have been defined for different purposes. The Bib-1 set is defined for bibliographic information while the Collections-1 attribute set is specifically for navigating digital collections.

Z39.50 has the advantage that the client does not have to know anything about the server, but it does raise the likelihood that results for the same query may vary quite widely: one server (Z39.50 target) may have a names index, another may have an authors index, another may not have a comparable index and may instead use a keyword index or produce an error. Standardisation of the settings for Z39.50 attributes can help, and a major initiative to achieve this was the establishment of the Bath Profile (ISO TC 46 SC 4, 2004) which regulates the Attribute types to help encourage interoperability.

In the distributed environment, Z39.50 is a valuable way of effectively creating a single resource, at least as far as the user is concerned. The data can continue to be held by those responsible for creating and maintaining it, but a central service can provide easy, unified access to any number of data sources.

Z39.50 can be criticised for being overly complex - functional sophistication tends to lead to inconsistent implementation. More lightweight solutions to distributed searching, such as SRU (Search and Retrieve via URL) may be seen as increasingly attractive options, as they make use of Web-based technologies. But Z39.50 is likely to have a lasting legacy: SRU builds upon Z39.50 concepts and semantics, which should enable the two standards to interoperate more effectively.
References
Z39.50 Maintenance Agency
http://www.loc.gov/z3950/agency/


Z39.50 on Wikipedia

http://www.ariadne.ac.uk/issue21/z3950/

The Bath Profile (ISO TC 46 SC 4, 2004)
http://www.collectionscanada.gc.ca/bath/tp-bath2-e.htm

Copac Z39.50 Interface: example of configuration information
http://copac.ac.uk/interfaces/z39.50/

EndNote bibliographic software
http://www.endnote.com/

Next month
Next month we will look at Search and Retrieve via URL (SRU), another standard for distributed database searching.

Jane Stevenson
Archives Hub, Mimas, University of Manchester