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
Search/Retrieve via URL

Name of Standards Developing Organisation
The Library of Congress

Current version
Version 1.2

Replaces
Version 1.1

Abstract
Search/Retrieve via URL (SRU) and Search/Retrieve via Web Services (SRW) are search protocols to query Internet indexes or databases remotely and return search results, specifying the syntax for both the search queries and the search results.

Description
The following URLs are searches for ‘roman baths’ using the syntax of three popular Internet search engines:


The queries are the same but the syntax implementing the queries is different. The responses not only contain search results but lots of formatting. SRU addresses these shortcomings by specifying the syntax for queries and results. SRU allows people and HTTP user agents to query Internet databases more seamlessly without the need of more expensive and complicated meta-search protocols.

SRU and SRW are known as Web Service-based protocols. A Web Service is a software system designed to support interoperable machine-to-machine interaction over a network. Web Services essentially send requests for information from a client to a server. The server reads the input, processes it, and returns the results as an XML stream back to the client.

There are two main types of Web service: REST-ful (Representational State Transfer) and SOAP-ful (originally Simple Object Access Protocol). SRW is a SOAP-
ful Web Service and SRU is a REST-ful Web service. Despite the differences in implementation, they are really very similar since they both define a similar set of commands (known as operations) and responses and they both support three operations: explain, scan, and searchRetrieve.

REST-ful Web Services encode the command from client to server in the query string of a URL and return a stream of XML data. The syntax of the query string and XML stream are dictated by the protocol. Because this is done within a URL, the communication is over an HTTP connection. SOAP-ful Web Services also make requests from a client to a server, but the XML documents are marked up in a SOAP XML vocabulary (this may also be called a SOAP envelope) which identifies the document as a SOAP message and includes call and response information. The response is also a SOAP XML document. SOAP requests and responses commonly use HTTP but they may use many other information transport mechanisms including email, Secure Shell (SSH) and telnet. SOAP is platform and language independent and provides a way to communicate between applications running on different operating systems, with different technologies and programming languages. REST-ful Web Services are generally seen as simpler to implement and requiring less overhead. SOAP is more robust and can be implemented in a larger number of networked environments.

SRU aims to reproduce the essential functions of Z39.50 in facilitating distributed searching and has essentially grown out of the search oriented Z39.50 protocol, as an attempt to bring it into line with current expectations and technology. SRU and OAI-PMH (which will be covered in the next issue of ARC) are complementary protocols. They have similar goals, namely the retrieval of metadata from remote hosts, but each provides functionality that the other does not. Both protocols have their own means to describe the properties of the remote service (SRU's explain operation and OAI-PMH's identify verb). Both protocols have a sort of 'browse' functionality (SRU has its scan function and OAI-PMH has ListSets). Scan is like browsing a book's back-of-the-book index; ListSets is similar to reading a book's table of contents.

There are marked differences between SRU and OAI when it comes to retrieval. SRU provides a richer set of query and response parameters, allowing a much more granular and precise approach. OAI-PMH is a useful means to allow a client to harvest the sum total of data a repository has to offer using a combination of the ListRecords and GetRecords verbs. This may be implemented at the expense of gathering a great deal of unwanted information. If a set of data were exposed via SRU as well as OAI-PMH, then SRU would have the advantage of allowing the client to identify and retrieve exactly the records required. OAI-PMH would be more appropriate if the client wanted to retrieve all data. It is also possible to combine the two approaches, using OAI-PMH to harvest and then using SRU as a search and retrieve mechanism.

SRU is not currently widely used in the archive community. The Archives Hub <www.archiveshub.ac.uk> is using SRU internally, as a means to gather data from distributed data providers (Spokes), as well as providing an external SRU interface to enable service providers to access the Hub descriptions remotely, maybe to provide their own federated search or to provide access to the descriptions from within their own environment. AdLib information systems is another example of a software system that has developed an implementation of SRU for its customers. This means that archives using AdLib can provide an SRU interface to enable others to search and retrieve the data remotely.
References

Search/Retrieves via URL Official Website
<http://www.loc.gov/standards/sru/>


Next month

Next month we will look at Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) in more detail than possible above.

Jane Stevenson
Archives Hub