

Z39.50

Overview of Z39.50

Z39.50 or ISO23950 is a protocol enabling search of and retrieval from remote databases. Its full name is ANSI Z39.50-1995, *Information Retrieval (Z39.50) Application Service Definition and Protocol Specification*. The body responsible for Z39.50 is ANSI/NISO. The standard initiative dates from 1984 with the Linked Systems Project that aimed to link the major library systems in the United States. The first edition appeared in 1988 and was replaced by the second edition in 1992. The 1992 edition broadened the scope of Z39.50 beyond library catalogue connectivity. The latest version was published in 1995.

The latest edition of Z39.50 was also put through the ISO fast track process to become ISO 23950 in 1998, replacing the Search and Retrieval (SR) standards ISO 10162 and 10163. Most implementations of SR, a protocol that was quite similar to Z39.50, were in Europe and the impetus to replace SR by Z39.50 came from the Europeans. SR was dropped in favour of Z39.50 as a move towards international interoperability.

This latest edition is being enlarged by the addition of new options namely by new search attributes, structure and content of records that are returned, clarifications and diagnostics concerning search behaviour. Additions and clarifications are maintained by the Z39.50 maintenance agency, reference: <http://lcweb.loc.gov/z3950/agency/>. An open membership international interest group, the Z39.50 Implementors' Group (ZIG) meets regularly to discuss additions and enhancements and is supplemented by a discussion list.

Adoption of Z39.50

The most common applications search remote databases that are connected to Internet, using the TCP/IP transportation protocol. This is not the only way in which Z39.50 is used, e.g. it can also be used to search CDROM databases in a local area network or a database on the same computer or network.

The protocol has been widely employed in library systems and has revolutionized the way in which library systems interact. The Library of Congress gateway (<http://lcweb.loc.gov/z3950/gateway.html#about>) lists over 400 library catalogues that are available representing for the large parts North American libraries or those from other parts who have used North American vendors for their implementation. Index Data in Denmark maintains a list also with just under 500 catalogues listed (<http://www.indexdata.dk/targettest/index.php>). The maintenance agency web page lists registered developers of which there are currently 75 (<http://lcweb.loc.gov/z3950/agency/register/entries.html>). To take into account many smaller implementations and those that have implemented the protocol behind firewalls or behind web based OPACs, the total number of libraries world wide that have made their catalogues publicly available using Z39.50 is likely to be well over

1000. Libraries are not the only implementations as the protocol has also been adopted by commercial databases, museums and government information centers.

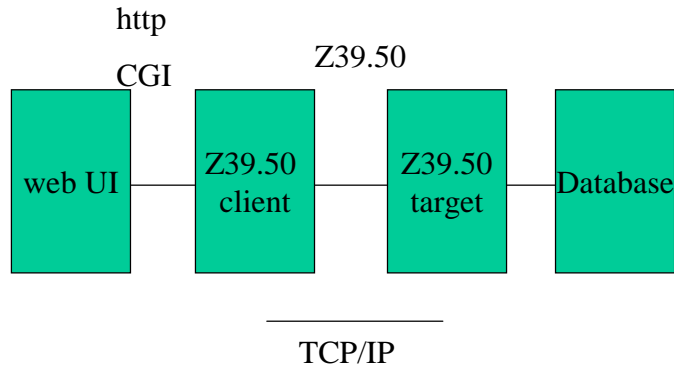
Major advantages

- It is possible to search multiple catalogues or databases with one single user interface
- It is possible to search multiple catalogues or databases with one search
- Results from multiple servers are in a standard format and can be combined. For example one of the various MARC formats is the most common format exchanged for library catalogue data.
- Results data can be reused by a client, e.g. for copy cataloguing or for creation of a bibliography
- Following from results, items can be ordered or requested on Inter-Library Loan

Z39.50 Architecture

Powerful clients that search multiple targets at once, convert and merge results for display and keep track of the progress of a query require disk space and memory to run and are commonly labeled “fat clients”. They also have a consequential cost. There has been a trend towards splitting clients into two with the user interface being world wide web enabled and capable of running using a web browser such as Internet Explorer or Netscape. The program part of the client, the fat part, is accessed from the web browser software and can be “shared” by many front end users thus giving a cost effective, widely accepted, modern configuration. The most common architecture installed is illustrated by the following diagram:

Most Common Architecture



The Essence of Z39.50

The main functions that Z39.50 provides are searching (search or browse) and retrieval. The client (or origin) can request the results in packet sizes that it can handle. Even long records can be split into two or more parts to facilitate transmission. For large results sets, the client can request that the server (or target) sorts the results before presentation so that the most relevant results (according to the sort criteria, e.g. most recent) are presented first.

Z39.50 messages

Table 1: Z39.50 Messages

Message from Origin	Response from Target	Comment
Initialisation Request	Initialisation Response	Establishes connection, can include password authentication, establishes options, e.g. character set used
Search Request	Search Response	Response can include: Diagnostic Results count plus first record Results count only
Present Request	Present Response	Follows on from a search request

Message from Origin	Response from Target	Comment
<i>No response</i>	Segment Request	Sent by the target to inform the origin that the results will be segmented
Delete Result Set Request	Delete Result Set Response	Enables deletion of a named result set. Not much used.
Sort Request	Sort Response	Request for the target to sort before results are presented.
Scan Request	Scan Response	Browse; designed to browse lines, index entries not records are given in the response
Access Control Response	Access Control Request	Sent by the target when authentication is required, e.g. to access restricted material
Resource Control Response	Resource Control Request	Sent by the target to warn about resource limits, e.g. getting low, results may not be complete
Trigger Resource Control request	<i>No response</i>	Sent by an origin to trigger a Resource Control Request from the target, e.g. to find out if resources are sufficient for a particular command
Resource Report Request	Resource Report Response	Sent by the origin requesting a report on resources
Extended Services Request	Extended Services Response	Sent by the origin to request a service, e.g. item order or update that may be performed by the target after the Z39.50 session has terminated
Close Request (or Response)	Close Response (or Request)	Ends the session. Either the target or the origin may send the request.

Z39.50 Session (Association)

When an origin links to a target, a session is established by the Initialisation request and response. The session is called a Z Association. A session may contain multiple messages, e.g. multiple queries but it can include a mixture of messages, e.g. search, scan and update. It is possible to relate a search to a previous search by using a feature called “named result set”. This feature enables refinement of searches and manipulation of result sets for presentation. The table below indicates an example of the sequence and types of messages that comprise a Z39.50 session or Z association.

Table 2: Example of Message Sequence in a Z Association

Message from Origin	Message from Target
Initialisation Request →	← Initialisation Response
Search Request →	← Search Response
Present Request →	← Present Response
Present Request (for more records or for same record/s, different content) →	← Segmentation Request
	← Present Response
Present Request (for next segment) →	← Present Response
Search Request (new or refining the previous request) →	← Search Response
Present Request →	← Access Control Request (e.g. for password)
Access Control Response →	← Present Response
Scan Request →	← Scan Response
Search Request (e.g. for a record behind an index entry from the Scan Response) →	← Search Response
Close Request →	← Close Response

Search and Present

Queries can be simple and can be more complex including the following features;

- Complex Boolean expressions (AND / OR / NOT and nesting)

- Proximity searching
- Comparison operators (equals, greater than etc.)
- Truncation

The Z39.50 search can be formulated in three ways. It can be in Reverse Polish Notation (RPN), in Common Command Language (CCL) or it can be in a proprietary query structure. It is mandatory for servers to be able to recognize the RPN structure. RPN queries consist of search attributes that may be combined with Boolean operators. There are various different sets of attributes, one for searching chemical data, another for searching holdings, another for searching cross domain, however the most common attribute set is Bib1 that is used for bibliographic searching. Bib 1 consists of 6 attribute types as follows:

Table 3: Bib 1 Attribute Set

Attribute Number and Name	Function, examples
1 Use	Indicates the Index or Access Point to search; e.g. title, author, subject etc.
2 Relation	Examples: equals, less than (e.g. for a date)
3 Position	Examples: First in field; first in subfield; anywhere
4 Structure	Examples: Word; phrase
5 Truncation	Examples: No truncation, right truncation; left and right
6 Completeness	Examples: Complete field; incomplete field

The Bib1 attribute set has grown in an uncontrolled way over the years. There is a proposal to replace it with two new attribute sets, Bib2 and the MARC attribute sets. The latter enables precise searching by MARC elements, e.g. a MARC tag and subfield.

Most Z39.50 messages are associated with search in one way or another, the core function of Z39.50 as illustrated in table two above. Initialise and Close relate to the Z association as a whole, while messages such as segmentation and access control could relate to search or to extended services.

Scan

Scan is a variant on search. It enables viewing of a sorted list, e.g. of indexed titles or authors and permits both forwards and backwards movement. Scan became a very popular way of searching to overcome the “opaqueness” of early searching systems that replied too often with “no matches found”.

There are many systems that support scan that can only support forwards movement largely because SQL, the access mechanism of many databases, only supports forwards browsing.

The results of scan are just indexed terms so that to retrieve a record from scan results, a follow on search is necessary. The follow on search from scan has recently been a focus of discussion, as a simple follow on search using the same term used for the original scan can produce unexpected results. Usually the results include more records using the term words in any position. A convention for conveying and reusing internal identifiers has been devised to solve this problem. This is within the Bath profile <http://www.nlc-bnc.ca/bath/bp-current.htm> .

Explain

Z39.50 includes an explain feature that is designed to allow a client to self configure once the TCP/IP and post address are known. It gives information on the searches a target accepts and the record formats that it delivers, together with various pieces of information regarding the many options such as character set negotiation etc.

There have not been many implementations of explain, although the European Union tried to encourage its uptake through many projects, including the Universe project that made an external database, accessible via Z39.50 explain that contained configuration and explain information for various Z39.50 targets. <http://www.fdgrouop.co.uk/research/universe> . This initiative also overcame the major limitation of Explain, that is the need to first discover the TCP/IP address and port number.

Client and Server or Origin and Target?

Z39.50 talks about “origin” and “target” rather than “client” and “server”. Why? The standard makes the distinction because there are many instances when one system can serve both roles. For example a Z39.50 system on a server may act as a “target” in relation to incoming searches but it may also act as a relay and forward searches to another “target”, thus acting as an “origin”.

Further, many systems these days have multiple levels as illustrated in diagram “Most common architecture”. In this example, the “client” in system terms consists of the web interface (the user interface) and the background program that constructs the Z39.50 queries. In Z39.50 terms, it is the second part that is the “origin”.

To add further confusion, in version 2 of the standard, and in the old version of the ISO standard 10162 and 10163, the term “source” was used instead of “origin”.

Z39.50 Systems compared with Web Search Engines

Web search engines index web pages. They do not index most databases that are accessible from web pages. Z39.50 enables searching of these databases and enables searches to be more precise than a web search. For example:

Search a name as an author, not as a subject

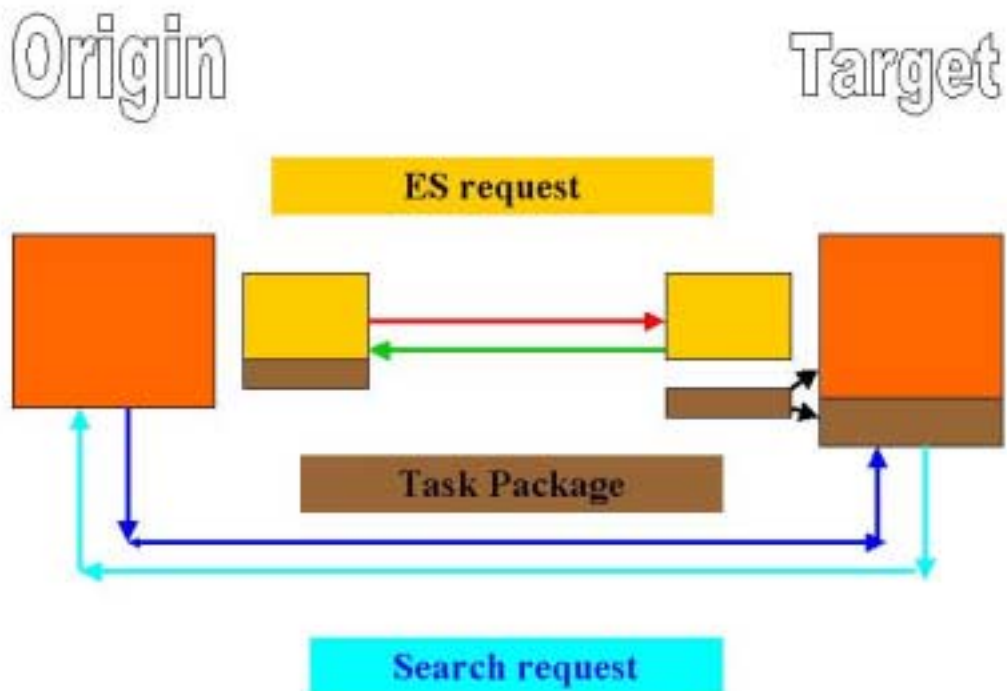
This is possible because Z39.50 includes definitions of common database access points such as author, title, ISBN etc. The protocol also provides for standardized delivery of search results. The client is able to stipulate both the format and the level of detail of the contents. The following formats (record syntaxes) are supported by Z39.50:

- MARC records (ISO 2709) (e.g. MARC21, UNIMARC, UKMARC)
- Simple displayable lines (SUTRS – simple unstructured transfer record syntax)
- Labeled data capable of display in the more recent versions of web browsers (XML)
- A generic self defining format (GRS-1 – Generic record syntax) – enables the transmission of groups of records, e.g. bibliographic record + holdings + images

Regarding content, the standard mandates a choice of brief or full records, but exactly what this means is at the discretion of the server. A more elaborate solution, eSpec, allows a client to request a customized version (e.g. full record but without notes) but this has not been widely implemented.

Extended Services – More than Just Searching

In addition to searching, Z39.50 has extended services that allow a client to send a command that may be performed or completed by the server after the Z39.50 searching session (or Z Association) has terminated.



Extended Services requests work by including two distinct parts in each request message. One part includes the request operation and the second part includes details of the task that is to be performed. By separating the two parts, the target is able to respond to the message then perform the task in its own timeframe. The results of the task are recorded in a special task database that is accessed using a normal Z39.50 search. There is a special Extended services attribute set for querying a task package database.

For example, an item order is sent as an extended services request. The target responds that the request has been received and is in process. When the item is despatched, to a ILL work station or via email directly to a user, the status of the task is updated in the task package database. To determine the status of the item order the Z39.50 origin sends a search request to the Extended Services database.

The target never initiates a transaction pair, with the exception of a security challenge, thus the origin always retains control and is responsible for all user interface. The origin may request interactive processing, otherwise the target chooses the scheduling of the process, whether background, interactive or batch.

The following table lists the services that are available via Z39.50 extended services.

Table 4: Extended Services

Extended Service	Comment
Persistent Result Set	Enables a set of search results to be saved on the server for later reuse.
Persistent Query	Enables a search to be saved for reuse at a later date, e.g. as the basis of a Selective Dissemination of Information (SDI) or Current Awareness Service (CAS).
Periodic Query Scheduler	Enables the scheduling of searches that were saved as persistent queries to be performed at regular intervals.
Item Order	Enables a request to be sent to the target for requesting an item usually from an external source, e.g. via inter-library loans.
Database Update	An origin can update a target database. Allows for insert, replace and delete.
Export Specification	A request to establish a record containing export (i.e. delivery) parameters, e.g. email system and address.
Export Invocation	A message requesting export (or delivery) of an item using a pre-defined export specification, e.g. a specific email address.

Update

The Union Catalogue Profile (UCP) <http://www.nla.gov.au/ucp> contains additional definitions for the maintenance of catalogues using Z39.50 update. The additional information provided in this profile includes:

- Update of Bibliographic, Authority and Holdings Data
- Control of concurrent update using a date/time stamp as a version identifier. Optional use of Record lock schema
- Specification of changes that can be sent as element update transactions (single field changes) and those to be sent as record replace transactions
- Merge and Global Change (batch search / replace) as special update
- Extensive diagnostics
- Conformance table - allowance for multiple conformance tables
- Sending both new and previous form of data (edit / replace record)
- Distributed review information

Update via Z39.50 offers the main advantage of providing feedback on the update process and the extensive diagnostics are a means of distributing the task of quality maintenance of a catalogue such as a union catalogue.

Innovative implementations

Implementing Z39.50 has proven advantageous in providing solutions to more problems than just remote database access. It has enabled systems of all ages, of various database models and schemas, and on different hardware and operating systems to become client server systems simply by superimposing a Z39.50 server program (the Z39.50 target) that reinterprets incoming queries and outgoing answers to be conformant with to the protocol.

Z39.50 is used in web OPACs as the means of querying the local library database. It has also been used in union catalogues as a means of switching between union catalogues and local catalogues. Other notable implementations include that of Endnote (see: <http://www.endnote.com/>), a client that searches databases with Z39.50 then permits the selection of records into a bibliography that can be stored on a personal computer.

Z39.50 has been behind several new initiatives, the most significant of which are:

- The Virtual Union Catalogue
- Cross Domain Virtual Catalogues (catalogues, databases, archives, museums)
- Copy Cataloguing and Real union catalogue maintenance

Virtual Union Catalogue

Using the ability to search several targets with a single search, a virtual union catalogue is created, obviating the overheads of maintaining a real union catalogue. The best known example is the Canadian union catalogue (VCUC) <http://www.nlc-bnc.ca/resource/vcuc/>. A lot of research has been directed at de-duplication of results. Most implementations have noted that multi-target searching is considerably slower, generally determined by the speed of the slowest server. The speed is made slower by sorting and de-duplicating and there is no compensation for lack of authority control that can fragment results and cause partial retrieval. Currently, there are limits to the effectiveness of sorting and de-duplication of large results sets, just when it is really needed. Interoperability problems leading to unpredictable results motivated the Bath profile initiative <http://www.nlc-bnc.ca/bath/bp-current.htm>.

Nevertheless, the virtual union catalogue remains an attractive proposition and is being used in Europe in the ONE projects <http://www.one-2.org/>.

Cross Domain Virtual Catalogues

Again, using the ability to search more than one target with one search, cross domain catalogues are emerging that search different types of target databases. The concept is an exciting one, e.g. allowing a search by a composer to retrieve:

- Books by and about the composer from a catalogue
- Manuscripts from an archive
- Articles by critics from a database
- Artifacts from a museum
- Photographs of buildings in which the composer resided from a local history database

To drive this, a new attribute set has been designed, the cross domain attribute set that is based on the Dublin Core data elements. The target will transform the cross domain attribute into a search appropriate for the database, e.g. “creator” may be interpreted as “author” for searching a catalogue but may be interpreted differently for searching a museum database. The Bath profile has determined that XML is a more appropriate format for returning results in response to a cross domain search.

Copy Cataloguing and Real Union Catalogue Maintenance

There has been an emergence of intelligent clients, particularly cataloguing clients that switch targets (servers) to produce efficient work flows, whilst at the same time presenting a single user interface. For example a cataloguing client can:

- Access library catalogues world wide as a source of copy cataloguing
- Search multiple targets from a single command input by the user either simultaneously or serially
- Access databases of authorized name and subject headings and code lists for selection and verification
- Update one or more databases with the catalogue record, e.g. local and regional or union catalogue

Z39.50 Version 4 2001

A new edition of Z39.50 is in progress. This new version will be largely a consolidation of minor amendments, clarifications and corrections that have been accumulating on the maintenance agency web pages. This is the natural maturing process that comes out of years of implementation experience. The changes are being made in a way that protects existing implementations in that there are additions but no substantial changes. The significant additions are as follows:

- encapsulation, enabling a group of transactions to be bundled together. This goes some way to making Z39.50 “stateless” and more suited for web based systems
- de-duplication request that enables an origin to request that duplicates be removed from a result set, usually after a sort request
- negotiation of character sets, language and record sizes
- optional query type using SQL

The new version is being produced following normal NISO review procedures. It will not stop the call from some members of the implementers’ community for a more radical set of changes that would modernize the standard but at the same time render existing implementations un-interoperable.

Advantages and Drawbacks of Z39.50

Z39.50 is a mature standard that has been widely implemented in certain communities. It could be argued that any database can be enabled to use the standard.

Nevertheless, the standard has not been widely accepted outside the library and its related community for several reasons.

- The standard is seen as complex and difficult to implement
- There are too many options, a fact that causes interoperability problems. It is possible to receive a reply that the server:
 - “cannot accept the query with those particular attributes or combination of attributes”
 - “cannot return records in the format required”
- It interoperates with the web but is not yet fully compatible according to some definitions

The ZIG (the Z39.50 Implementers’ Group) is making attempts to address these issues.

Difficult to implement

The standard is not easy bed time reading, so it is difficult to implement just using the standard alone. This has been considerably alleviated by the provision of tool kits, test clients and test servers. The Z39.50 maintenance agency (<http://lcweb.loc.gov/z3950/agency>) lists up to date sources of these development and testing aids. In addition the ZIG list is very active with several hundred implementers available world wide to field technical questions. The ZIG has considered making an implementers’ guide.

Configuring targets for Z39.50 applications is the next area focused for improvement. To configure a target, it is necessary to know:

- The TCP/IP address and port number
- How to search the target – which searches it supports and which search attributes
- How to retrieve – which record syntaxes and schemas are supported

Z39.50 includes an explain feature that is designed to allow a client to self configure once the TCP/IP and post address are known. This has not been a success and in practice most configurations are set manually. Widespread adoption of standard profiles will considerably simplify the configuration process as well as improving interoperability. Discussions are now centered around creating a directory of Z39.50 targets using UDDI

<http://lcweb.loc.gov/z3950/agency/zig/meetings/dc2000/presentations/uddi.ppt> or an updated version of the ISO directory standard ISO 2146.

Interoperability Problems

Profiles are emerging that contain standard settings for queries and responses. The main profiles are:

Domain	Recognized Profile
Libraries	Bath http://www.nlc-bnc.ca/bath/bp-current.htm
Museums	CIMI http://www.cimi.org/publications.html#z3950_2
Government Info	GILS http://www.gils.net/prof_v2.html
Thesauri	Zthes http://lcweb.loc.gov/z3950/agency/profiles/zthes-04.html
Geospatial databases	GEO http://www.blueangeltch.com/standards/GeoProfile/geo22.htm

Recently, the Bath profile has attracted considerable attention and is gaining support. The profile is ambitious in tackling the interoperability issues and provides for staged implementation. Notably, it includes:

- Searching bibliographic databases
- Searching holdings databases
- Searching cross domain

Web compatibility

There is not yet a consensus but there is much discussion on making the protocol more elegantly coupled with the web, for example by being XML based using eye readable text rather than BER that uses compiled coded data. There is no longer a strong argument for restricting the amount of data passed as the amount of space saved by encoding a search command is insignificant compared with the size of the average image exchanged in gif or jpeg format.

One of the main areas of incompatibility with the web is in the area of “statefulness”. Z39.50 currently operates in sessions. A connection is made with the database and multiple queries can be made until the session (Z association) is closed. This allows queries to be evolved, e.g. by combining results sets or refining them by including them in a query. Nevertheless, this is not the way the web operates. Each time a page is returned to a web browser, the process is terminated. Connection (initialization and authentication) and query are sent in the one command.

To eliminate statefulness, the ZIG has already adopted encapsulation, but is also discussing more simple measures that would change the standard such as making initialization optional and allowing search commands to include sort. [One project, tentatively named ZNG \(Z39.50 Next Generation\), is testing search and retrieve using just the HTTP protocol. The messages can be sent either as simple URLs \(uniform resource locaters which are the strings that are typed in the address input section of the web browser\) or, if more complicated, as XML documents within a SOAP wrapper. The responses are always sent as XML documents within a SOAP wrapper.](#)

The messages are eye readable, including simple structures such as “bath_ti_kw = brideshead” meaning Bath keyword title. By retaining the important concepts of Z39.50, but modernizing and simplifying the communication techniques, the testing partners are hoping to attract implementers such as online bookshops who have up until now stayed shy of the standard. Pica is one of the testing partners in this project.

The Contenders

XML Query

XML query (previously provisionally called XQL) is an emerging standard being developed by the W3C. It is heavily influenced by SQL, a standard way of searching databases. XML query is not a full alternative to Z39.50 because it handles searching but not the management of results. Moreover, XML Query lacks a list of commonly understood access points such as title, therefore before querying a database it is necessary to discover its structure. This makes multi-target searching difficult.

Because of the weight of W3C and SQL, XML query is likely to be adopted. We will see web browsers accept “xmlq://” as well as “http://” but we are less likely to see “z3950://”

It is possible that Z39.50 could encompass XML query by allowing it as an alternative query type.

See: <http://www.loc.gov/z3950/agency/zig/meetings/dc2000/zxml-report.html>

Open URL

Open URL offers a means of searching a remote database using a standard identifier such as ISBN, DOI, etc. Although it is far simpler and more restricted than Z39.50, it will probably coexist in Z39.50 systems and be the choice for searching via identifier. Where an identifier is not known for a known item, Z39.50 could be used, e.g. searching via author and title.

Conclusion

Neither of these contenders is poised to fully replace Z39.50, therefore Z39.50 is not likely to die in the near future. What Z39.50 is facing, is the proposition of being relegated to a backwater and for the important concepts that it has devised and realized being ignored by mainstream developments. This can be avoided if the protocol is modernized to run easily on existing platforms and to be based on related main stream standards and tools. A good discussion of this issue is by Bill Moen *Resource Discovery using Z39.50: Promise and Reality* January 2001 http://www.loc.gov/catdir/bibcontrol/moen_paper.html .

References

ANSI Z39.50-1995, *Information Retrieval (Z39.50) Application Service Definition and Protocol Specification*. NISO 1995. (Available for download from <http://lcweb.loc.gov/z3950/agency/>)

Application Profile for the Government Information Locator Service (GILS)". Available at: http://www.gils.net/prof_v2.html

The Bath Profile: An International Z39.50 Specification for Library Applications and Resource Discovery. Release 1.1 Internationally Registered Profile. June 2000. Available at: <http://www.nlc-bnc.ca/bath/bp-current.htm>.

CIMI Z39.50 Profile. Available at: http://www.cimi.org/publications.html#z3950_2

Endnote: bibliographies made easy. Available at: <http://www.endnote.com/>

ONE2 EU-project in DG XIII, 4th Framework. Available at: <http://www.one-2.org/>

Resource Discovery using Z39.50: Promise and Reality Bill Moen, January 2001 http://www.loc.gov/catdir/bibcontrol/moen_paper.html .

Standardized Pica3 protocol, Clemens Buijs, Pica 1996.

Union Catalogue Profile: Internationally Registered Profile to be used with ISO 23950 (Z39.50 Information Retrieval Protocol) Extended Service for Update, June 1999. Available at: <http://www.nla.gov.au/ucp>.

The UNiVerse Project. Available at: <http://www.fdgrouop.co.uk/research/universe>.

The Virtual Canadian Union Catalogue Project. Available at: <http://www.nlc-bnc.ca/resource/vcuc/>

Z39.50 Application Profile for Geospatial Metadata or "GEO". Available at: <http://www.blueangeltech.com/standards/GeoProfile/geo22.htm>

Z39.50 as an XML Protocol. Available at: <http://www.loc.gov/z3950/agency/zip/meetings/dc2000/zxml-report.html>

Z39.50 register of implementers. Z39.50 Maintenance agency. Available at: <http://lcweb.loc.gov/z3950/agency/register/entries.html>

The Z39.50 target directory. Copenhagen: Index Data. Available at : <http://www.indexdata.dk/targettest/index.php>.

Zthes: a Z39.50 Profile for Thesaurus Navigation: Available at: <http://lcweb.loc.gov/z3950/agency/profiles/zthes-04.html>