**Z39.50 for All**

Paul Miller looks at the Z39.50 standard and attempts to extract some meaning from the mass of associated literature.

---

**Introduction**

Z39.50. Despite certain nominative similarities, it's not a robot from that other blockbuster of the summer, *Star Wars: The Phantom Menace*, but rather the cuddly and approachable name for an important standard of relevance to many working with information resources in a distributed environment. In this particular summer blockbuster (*Ariadne*, to which I'm sure many readers frequently refer in the same paragraph as *Star Wars*), I'll attempt to remove some of the mystique surrounding this much-maligned standard, and illustrate some of what it can be used for.

**Z39.50 quick facts**


The formal home of the standard is the Z39.50 Maintenance Agency [2], hosted by the United States' Library of Congress [3]. Continued development takes place within an informal group of implementors and developers known as the Z39.50 Implementors Group, or ZIG. The work of the ZIG is progressed on an active mailing list [4], and through two or three face-to-face meetings each year, the most recent of which was held in Stockholm in August 1999. These meetings are open to all, with the next one scheduled for January 2000 in San Antonio, Texas [5].

Despite a common misconception to the contrary, Z39.50 is not simply used by libraries, although the library sector is one with a clear and long-held need for Z39.50-type functionality. In the cultural heritage sector, for example, the Consortium for the Computer Interchange of Museum Information (CIMI) worked to develop a Profile meeting the needs of cultural heritage practitioners [6]. In the world of government and community information, too, the Government Information Locator Service (GILS) Profile [7] makes use of Z39.50 to link a wide range of resources internationally.

There is a huge amount of information available on the web relating to Z39.50. A basic search for the term on 6 September produced a daunting 2,863 hits from Alta Vista [8], a scary 23,002 from Northern Light [9], and a positively mind-numbing 27,651 from FAST [10]. See Phil Bradley's Search Engine article [11] in this issue for some more information on these search engines. It is, of course, difficult to navigate meaningfully through such a mountain of information, but two sites repeatedly prove useful first stops. These are the Maintenance Agency's own web site [2], and Dan Brickley's Z39.50 Resources page at ILRT [12]. *Biblio Tech Review*’s technical briefing on Z39.50 [13] is also valuable, especially as an overview.

**What does Z39.50 do?**

Z39.50 is designed to enable communication between computer systems such as those used to manage library catalogues. This communication could be between a cataloguer's PC (or an OPAC terminal in public use) and the
library catalogue itself, running on a Unix server in the basement, and equally it could be between a user browsing the web from Hull (East Yorkshire), a library OPAC in Hull (Quebec), a GILS database in Hull (Florida), and a museum collection management system in Hull (Texas)! Although the former is undeniably useful to those individuals who have to update things such as library catalogues, it is the latter application and others like it which represents much of the potential of Z39.50 in today's distributed network environment.

Real examples of this type of distributed search include the Gateway to the Arts & Humanities Data Service [14] in the UK, and the Melvyl® system of the University of California [15].

**Arts & Humanities Data Service**

The AHDS Gateway [16], physically based in London, uses Z39.50 to query five totally different databases containing information on archaeology (York), history (Colchester), the performing arts (Glasgow), the visual arts (Newcastle), and textual studies (Oxford). The databases describe different data types according to different cataloguing standards. They are driven by different database management software and run on a variety of hardware platforms. Nevertheless, despite these differences, the combination of Z39.50 and Dublin Core elements is sufficient to enable meaningful searches across the five sites.

**AHDS Gateway: Advanced Search**

Use this form to conduct more advanced searches of Service Provider catalogues and other resources. To issue a simple keyword search of Service Provider catalogues use the Simple Search page.

**Select AHDS Catalogues:** Click on the boxes below to include a catalogue in your search. A tick in the box indicates a catalogue is included.

- Archaeology Data Service
- History Data Service
- Oxford Text Archive
- Performing Arts Data Service
- Visual Arts Data Service (including ADAM)

Choose a field name to search (e.g., all, subject), and a keyword or expression to search for (e.g., "city", "Impressionism", "Dickens"). Use the operators (And, Or, and And Not) to construct more complex queries. See help for further details.

**Figure 1: Arts & Humanities Data Service Gateway - Advanced Search**

The results of a query across the services are summarized and presented to the user as shown in Figure 2. The status bar at the bottom of this figure illustrates progress of the query against each of the five targets. In the case shown here, both the Archaeology Data Service and Performing Arts Data Service have yet to respond.
Melvyl®

The University of California's Melvyl® system allows users to submit searches across the distributed libraries of the University of California (University of California at Los Angeles, University of California at Santa Barbara, University of California at Irvine, etc.). Melvyl® is just one resource within the California Digital Library [17], which describes itself as

...(the) tenth library for the University of California (UC). A collaborative effort of the UC campuses, organizationally housed at the University of California Office of the President, it is responsible for the design, creation, and implementation of systems that support the shared collections of the University of California.

http://192.35.215.185/mw/mw.cgi?sesid=0904727006&Cecho(home/intro)&Z39.50_for_All,_Ariadne_Issue_21

http://www.ariadne.ac.uk/issue21/z3950/
Figure 3: The California Digital Library
Where is it hiding?

Both Melvyl® and the AHDS Gateway make use of Z39.50 in order to create a single, virtual, resource. In both cases, the data themselves remain distributed and under the control of those with the knowledge and expertise to most effectively maintain, update, and add to the resources already present. By using Z39.50, these obvious benefits of distributed data management are combined with the equally valuable benefits of unified data access, which allows the user to submit a single search across multiple resources, regardless of their physical proximity to one another or to the user.

Like a large number of other applications of Z39.50, the technology is effectively hidden from the user in both these examples; despite what the informational screens on the web sites might say, so far as the user is concerned they're simply searching one great big database. Z39.50 also crops up behind a large number of other services with which readers may be familiar, although some are possibly more unexpected than others. The United Kingdom's JISC-funded Mirror Service [18], for example, makes use of Z39.50 internally in managing the services that it mirrors. This increasing invisibility of Z39.50 is all very well for the end user, but makes it difficult to sell to people who will all too often claim that Z39.50 isn't used any more. It most definitely is. Developers have simply become good — possibly too good, given their need to raise awareness in order to sell future products? — at hiding it.

How does it do it?
Everyone following along OK so far? Well, unfortunately this is where things get a little — but hopefully not too — complicated...

Clients and Servers versus Origins and Targets

Z39.50 follows what is known in Computing as a client/server model, where one computer (the client or, in Z39.50 parlance, the 'Origin') submits a request to another computer (the server or, to Z39.50, the 'Target') which then services the request and returns some kind of answer. As we saw above with queries being sent to multiple databases simultaneously, there can be more than one server/target, although the user will normally only be sat in front of one client/origin at any given time.

Facilities and Services

Z39.50 is divided into eleven basic structural blocks, known as Facilities. These Facilities comprise Initialization, Search, Retrieval, Result-set-delete, Browse, Sort, Access Control, Accounting/Resource Control, Explain, Extended Services, and Termination. Each Facility is divided into one or more Services, and it is these Services that people usually talk about. A Service facilitates a particular type of operation between the Origin and the Target, and Z39.50 applications select those Services which will be needed in order for them to fulfil their function. Of all the Services, the three most basic are Initialization (Init), Search, and Present, and all of these should be found in the majority of Z39.50 applications.

Init is the first step in any query process, and involves the Origin making itself known to the Target, and agreeing a few 'ground rules' for the manner in which subsequent queries will be handled. Systems which require passwords will also exchange authentication details at this stage.

Search is where the majority of the work is done, as it is this Service which enables the Origin to submit queries to the Target. These queries may range from the very simple right up to quite complex Boolean queries (AND, OR, NOT, >, <, etc.). Although Profiles (see below) often make statements about the Services to be supported by any conformant application, the bulk of many Profiles is given over to specifying the attribute combinations for meaningful use of Search.

Present is used to control the manner in which results are returned to the user. Within Present, a user could ask for the first ten records of a large result set, or request that the data be returned in a different Record Syntax; UKMARC instead of USMARC, for example.

Although these are the main Services, there are a further ten, summarized below. More detail may be found in Section 3.2 of the ANSI/NISO document for Z39.50 version 3 [1].

<table>
<thead>
<tr>
<th>Facility</th>
<th>Service</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initialization</td>
<td>Init</td>
</tr>
<tr>
<td>Search</td>
<td>Search</td>
</tr>
<tr>
<td>Retrieval</td>
<td>Present</td>
</tr>
<tr>
<td>Retrieval</td>
<td>Segment</td>
</tr>
<tr>
<td>Result-set-delete</td>
<td>Delete</td>
</tr>
<tr>
<td>Browse</td>
<td>Scan</td>
</tr>
<tr>
<td>Sort</td>
<td>Sort</td>
</tr>
<tr>
<td>Access Control</td>
<td>Access-control</td>
</tr>
<tr>
<td>Accounting/Resource Control</td>
<td>Resource-control</td>
</tr>
<tr>
<td></td>
<td>Trigger-resource-control</td>
</tr>
<tr>
<td></td>
<td>Resource-report</td>
</tr>
<tr>
<td>Explain</td>
<td>uses the Services of Search and Retrieval</td>
</tr>
<tr>
<td>Extended Services</td>
<td>Extended-services</td>
</tr>
</tbody>
</table>
Pardon?

OK, let's try explaining it with an example...

Simplifying hugely, Init might be seen as a greeting from the Origin ("Hello, do you speak English?") and a related response from the Target ("Hello. Yes, I do. Let's talk"). Without this positive two-way dialogue, the session cannot proceed.

A Search request is then transmitted from the Origin ("OK — can I have everything you've got about a place called 'Bath'"), and is responded to by the Target ("I've got 25 records matching your request, and here are the first five. As you didn't specify anything else, I've sent them to you in MARC format, so I hope that's OK").

Finally, the Origin uses Present to ask for the data they want ("25, eh? Can I have the first ten, please. Oh, and I don't really like MARC. If you can send me some Dublin Core that would be great, and if not I'll settle for some unstructured text (SUTRS)"), resulting in the transmission of the records themselves from the Target.

Ah... Now I understand...

That's all very well, and probably even appears pretty straightforward. However, there's just a little more that needs to be covered before you can join the ranks of the Z-cognoscenti.

Attribute Sets

The manner in which the Search process is governed is closely linked to the Attribute Sets being used by both Origin and Target. An Attribute Set is, as the name suggests, a set of attributes. Each of these attributes can have one of several values, and these values govern the manner in which a search proceeds.

Perhaps the best known Attribute Set is Bib-1 [19], an Attribute Set originally designed for bibliographic resources but now commonly used for a wide range of applications.

Bib-1 comprises six groupings of attributes, or Attribute Types. These are

- **Use Attributes**, which define the access points for a search (title, author, subject, etc.)
- **Relation Attributes**, which determine how the search term entered by the user relates to values stored in the database index (less than, greater than, equal to, phonetically matched, etc.)
- **Truncation Attributes**, which define which part of the value stored in an index is to be searched (the beginning of any word in the field, the end of any word in the field, etc.). i.e. on a search for 'Smith', does the stored value start with 'Smith', end with 'Smith', or something else entirely?
- **Completeness Attributes**, which specify whether or not a search term can be the only value in an index.
- **Position Attributes**, which specify where in an index field the search term should occur (at the beginning, anywhere, etc.)
- **Structure Attributes**, which specify the form to be searched for (a word, a phrase, a date, etc.)

Attention normally only focuses upon discussion of **Use Attributes**, but it is important to correctly set values for the other five attribute types as well if a search is to have maximum value. These other attribute types are frequently interpreted differently by the suppliers and specifiers of Z39.50 systems, making it difficult for users to interoperate with 'unknown' systems from a vendor with which they are unfamiliar. Work on Profiles such as the Bath Profile (see below) has gone some way towards standardizing the setting of all these attribute types, thus increasing the likelihood of reliable searches across a wide range of Z39.50 systems.

Record Syntaxes

Once the user has undertaken a search, and discovered that there are a number of records meeting their search criteria, the next decision to be made is how the data are transmitted back for display. Z39.50 has a notion of Record Syntaxes, and the Origin is able to request that data be transmitted in one of these Syntaxes. If the Target is able to comply, it will do so. If it cannot, it is possible for the Target to send the records back in some other format that it is capable of.
Given the widespread adoption of Z39.50 in the library world, almost every Z39.50 Origin or Target is capable of handling at least one of the flavours of MARC. USMARC is commonly supported, often with regional support for UKMARC, DANMARC, AUSMARC and others, depending upon where the vendor is based, and who they see as their key market.

Z39.50 is not simply used for transferring MARC records, and other Record Syntaxes exist to meet a variety of other requirements. These syntaxes include SUTRS (Simple Unstructured Text Record Syntax; a raw ASCII text file, lacking in any formatting or structure), GRS-1 (Generic Record Syntax; a flexible structure within which almost any database structure can be replicated), and XML (the World Wide Web Consortium's eXtensible Markup Language; another flexible structure, and one with which web applications and developers will be familiar).

**Profiles**

Profiles in Z39.50 tend to be used in order to gather a particular suite of Attributes, Record Syntaxes, and other factors together in order to meet the needs of a particular community, whether that be subject, area, or application based.

As can be seen from the Maintenance Agency's list [20], Profiles span a wide range of task areas, including a Profile for the Geospatial community, one for Government Information, one for the Cultural Heritage sector, and others. These Profiles are often developed within the community with a requirement for them, with the ZIG and the Maintenance Agency serving in an advisory and support role.

**Towards the future: the new Attribute Architecture**

As mentioned above, the Bib-1 Attribute Set was originally intended for use in searches for bibliographic resources such as books. Over the years, however, other communities have increased their use of Z39.50, and have tended to add new Use Attributes to Bib-1, rather than creating wholly new Attribute Sets. This gradual accretion of Use Attributes into Bib-1, together with a number of overlaps between this Attribute Set and those few which had been created by other communities led members of the ZIG to realize that a restructuring was called for before matters became any more complex.

The new Attribute Architecture [21] attempts to solve these, and other, problems by recognizing that certain collections of Attributes are likely to be common to many applications, whilst others will be quite specialist and specific. It is intended that the commonly used attributes should be collected together in new core Attribute Sets which can be used by all applications, regardless of the area in which they are focused. These applications can then add functionality from localized and domain-specific Attribute Sets without duplicating the common functions offered by the core sets.

Along with other improvements, the new Attribute Architecture changes the Attribute Types familiar from Bib-1;

<table>
<thead>
<tr>
<th>New Attribute Type</th>
<th>Roughly corresponding Type from Bib-1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Access Point</td>
<td>Use</td>
</tr>
<tr>
<td>Semantic Qualifier</td>
<td>new Attribute Type</td>
</tr>
<tr>
<td>Language</td>
<td>new Attribute Type</td>
</tr>
<tr>
<td>Content Authority</td>
<td>new Attribute Type</td>
</tr>
<tr>
<td>Expansion/ Interpretation</td>
<td>Truncation and some of Relation</td>
</tr>
<tr>
<td>Normalized Weight</td>
<td>new Attribute Type</td>
</tr>
<tr>
<td>Hit Count</td>
<td>new Attribute Type</td>
</tr>
<tr>
<td>Comparison</td>
<td>most of Relation and part of Completeness</td>
</tr>
<tr>
<td>Format/ Structure</td>
<td>Structure</td>
</tr>
<tr>
<td>Occurrence</td>
<td>Completeness (more or less)</td>
</tr>
<tr>
<td>Indirection</td>
<td>new Attribute Type</td>
</tr>
</tbody>
</table>
A lot of work remains to be done on the Attribute Architecture, but two key components are now more or less complete. These are the Utility Attribute Set [22], which relates to the description of records, and the Cross Domain Attribute Set [23], which relates to the description of the resources those records describe.

Despite ongoing work by a number of vendors (e.g. Hammer, pers comm.), it seems likely that the new Attribute Architecture will take a number of years to establish itself and to become fully integrated within a new generation of tools.

**Z39.50 and the DNER**

Given the range of functions for which it was designed, Z39.50 is likely to play an important role in the initial phases of the Distributed National Electronic Resource (DNER) being constructed in the UK by the Joint Information Systems Committee (JISC).

This DNER is seeking to link existing services for the Higher Education Community (such as AHDS, the Data Centres at Edinburgh, Bath, and Manchester, the Electronic Libraries' Programme Clumps and Hybrid Libraries, etc.) in order that users both gain a greater awareness of the range of resources available to them and gain enhanced access to the contents of those resources. The DNER will also seek to provide access to resources from beyond the UK Higher Education community.

The resources to be linked by the DNER are highly diverse, and Z39.50 currently appears to be the only viable means by which access to as many of them as possible can be achieved. Under this model, the individual resources would be associated with a Z39.50 Target. Origins such as the one already in place for the AHDS [16] could be established at a variety of levels, perhaps including nationally (a JISC Gateway to its content), by faculty (the AHDS model), by data type (a Gateway to moving image data), and locally (the University of Hull's Gateway to services it has bought access to, whether within the DNER or independently from suppliers such as SilverPlatter).

As has been mentioned, above, there is a wide range of Profiles available to cover some of the services making up the DNER. There is also room for interpretation within and between these Profiles as to how certain Attribute Types are best handled for particular queries. Such ambiguity makes a model like that proposed for the DNER extremely difficult to implement without a high degree of communication and control between participants. An evolving development known as the Bath Profile, however, offers a means by which the DNER's needs can be met; every conformant Origin and Target will, to a certain degree, be visible to every other conformant Origin and Target, and queries will be formulated and responded to in a standardized manner with the result that users can place greater reliance upon the answers being returned. Adoption of the Bath Profile does not prevent Origins and Targets also making use of other Profiles more closely related to their particular community or data form.

**The Bath Profile**

This Profile, currently known as *The Bath Profile: An International Z39.50 Specification for Library Applications and Resource Discovery* has been evolving for some time under the guidance of Carrol Lunau at the National Library of Canada.

It specifically addresses the need for a relatively simple Profile in which the values of all Attribute Types are closely regulated, such that all conformant systems may be expected to behave in a particular manner in response to a small number of defined query types. The Profile is suitable both for library applications and for searches across a number of non-bibliographic domains.

Long-running discussion on the ZIP-PIZ-L mailing list and in a series of teleconferences culminated in a meeting held in Bath during August of 1999 [24]. At this meeting, the authors of a range of existing Profiles gathered along with a number of vendor representatives and finalised their ideas on the Profile. It is currently expected that a draft of the Profile will be made widely available for comment towards the end of September 1999, and this draft will be announced on various electronic mailing lists once it is ready, as well as being linked to the meeting web site [24]. A number of those present at the meeting have already stated that the Profiles for which they are responsible will be modified to align directly with the Bath Profile. In these cases, the Bath Profile will become the core of a larger Profile which will also include extra functionality to meet local requirements.

**Conclusion**
Z39.50 is often attacked on a variety of levels by those who see it as overly complex, old fashioned, not sufficiently web-like, or simply no match for the latest 'great idea' (currently, this 'great idea' is usually cited as being some combination of XML and RDF). It is undoubtedly true that Z39.50 has quirks and limitations, some of which have been outlined in this paper. This is true, though, of most standards, and the very fact that Z39.50 has been extensively used for long enough to be criticized as old fashioned is surely a testament both to its robustness and to the lack of any viable alternative.

New technologies such as XML and RDF certainly fulfil aspects of the information discovery and retrieval process better than basic Z39.50, but work is underway to capitalize upon this, and to tie such technologies more closely to Z39.50. It appears for the moment that, whatever its limitations, Z39.50 remains the only effective means of enabling simultaneous queries upon distributed heterogeneous databases, and this remains something that the broader user community wants to be able to do.

**Glossary**

A number of the principal terms and abbreviations used in this paper are defined, below.

**Attribute Set**
A collection of Attribute Types (Use, Relation, Completeness, etc.), gathered together to address a particular purpose. The best known Attribute Set is Bib-1.

**DNER**
The Distributed National Electronic Resource. The DNER is being built by the Joint Information Systems Committee (JISC) of the Higher Education Funding Councils in the United Kingdom, and will link users to the wealth of resources provided for them by JISC-funded Data Centres and others.

**GILS**
The Government Information Locator Service. This was developed in the United States as a distributed collection of 'gateways' providing public access to information from government. GILS has since been widely adopted in other countries around the world, and a closely related *Global* Information Locator Service was initiated under the auspices of the G7 countries. The GILS Profile [7] for Z39.50 is one way in which these distributed gateways may be searched.

**Origin**
A piece of software responsible for submitting a user request to one or more Targets. The term might also be applied to the user him/herself or to the computer at which they are sitting. More commonly known outside the Z39.50 community as a 'client'.

**Profile**
A specific interpretation of the manner in which Z39.50 — or a subset thereof — should be used to meet the needs of a particular *application* (the GILS Profile for the Government Information Locator Service), *function* (ATS-1 for author/title/subject searches), *community* (the CIMI Profile for cultural heritage), or *environment* (the CENL Profile for European National Libraries).

**RDF**

**Service**
One of the basic building blocks of Z39.50. These Services include Init(ialisation), Search, Present, Explain, etc.

**Target**
A piece of software responsible for passing requests from an Origin to the database on top of which it sits. The term might also be applied to the database itself or to the computer on which it resides. More commonly known outside the Z39.50 community as a 'server'.

**Use Attribute**
A Use Attribute specifies an 'access point' onto the underlying database with which a Target is associated. Use Attributes include such things as 'title', 'author', 'subject', etc., and they are normally mapped onto similar fields in the underlying database.

**XML**

**Z39.50**

**ZIG**
The Z39.50 Implementors Group, an informal body of suppliers and developers through which continued development of Z39.50 takes place.

References

   http://lcweb.loc.gov/z3950/agency/1995doce.html
2. Z39.50 Maintenance Agency
   http://lcweb.loc.gov/z3950/agency/
3. Library of Congress
   http://lcweb.loc.gov/
4. How to subscribe to the ZIG listserv
   http://lcweb.loc.gov/z3950/agency/zig/listserv.html
5. Future ZIG meetings
   http://lcweb.loc.gov/z3950/agency/zig/meetings.html
6. A Z39.50 Profile for Cultural Heritage Information
   http://www.cimi.org/standards/index.html#THREE
7. Application Profile for the Government Information Locator Service (GILS), version 2
   http://www.gils.net/prof_v2.html
8. Alta Vista
   http://www.altavista.com/
9. Northern Light
   http://www.northernlight.com/
10. FAST Search
    http://www.alltheweb.com/
11. Phil Bradley, FAST - the biggest and best yet? Ariadne, No. 21, September 1999
    http://www.ariadne.ac.uk/issue21/search-engines/
12. Dan Brickley, Z39.50 Resources
    http://www.ilrt.bris.ac.uk/discovery/z3950/resources/
13. Z39.50, Biblio Tech Review
14. Arts and Humanities Data Service
    http://ahds.ac.uk/
15. Melvyl
    http://www.melvyl.ucop.edu/
16. AHDS Gateway
    http://prospero.ahds.ac.uk:8080/ahds_live/
17. California Digital Library
    http://www.cdlib.org/
18. UK Mirror Service
    http://www.mirror.ac.uk/
19. Bib-1 Attribute Set
    http://lcweb.loc.gov/z3950/agency/defns/bib1.html
20. Z39.50 Profiles
    http://lcweb.loc.gov/z3950/agency/profiles/profiles.html
    http://lcweb.loc.gov/z3950/agency/attrarch/attrarch.html
22. Z39.50 Utility Attribute Set, Draft 3
    http://lcweb.loc.gov/z3950/agency/attrarch/util.html
23. The Z39.50 Cross-Domain Attribute Set, version 1.4
    http://www.oclc.org/~levan/docs/crossdomainattributeset.html
24. http://www.ukoln.ac.uk/interop-focus/activities/z3950/int_profile/bath/

Author Details

Paul Miller
Interoperability Focus
UKOLN
Acknowledgements

UKOLN is funded by the Library and Information Commission, the Joint Information Systems Committee of the Higher Education Funding Councils, as well as by project funding from the JISC and the European Union. UKOLN also receives support from the University of Bath where it is based.

Interoperability Focus is based at the University of Hull, and receives additional support from this institution.