# Joined up writing: an Internet portal for research into the Historic Environment

Tony Austin[1], Francisco Pinto[2], Julian Richards[1] Nick Ryan[2]

*1. Archaeology Data Service, Department of Archaeology, The King's Manor, University of York, York, YO1 7EP, UK.*
*2. Computing Laboratory, University of Kent at Canterbury, Kent, CT2 7NF, UK.*

## Abstract

Resource discovery has been a major concern of the Archaeology Data Service (ADS) since its inception in 1996. The ADS was heavily involved in the promotion and adoption of the Dublin Core (DC) metadata standard for categorising resources and how this can facilitate interoperability between distinct resources. An early example of this was the Z39.50 enabled Gateway which enables the cross searching of the metadata indexes held by the various service providers (including the ADS) that comprise the Arts and Humanities Data Service (AHDS).

Research into the Historic Environment is particularly concerned with spatial and chronological attributes of data. Dublin Core provides for this with its Coverage element. Chronological descriptors such as thesaurus derived period terms or a specific date can be referenced as can locative information such as site, place, unitary authority and country. More specific geospatial information in the form of coordinates from a grid referencing system can also be recorded. The ADS provides access to some 400,000 indexing records about the Historic Environment nearly all of which are thus referenced.

Coordinate referencing has allowed the development of various map-based search options for interfacing with ADS resources; initially relying on user input of coordinates to define a search area but more recently using click-on maps.

A recent project with the Computer Laboratory of the University of Kent at Canterbury (UKC) has successfully drawn these various strands together for the Z39.50 enabled searching of a number of geographically remote datasets. Other partners in this project include the Portable Antiquities Scheme (PAS), the Royal Commission on the Ancient and Historic Monuments of Scotland (RCAHMS) and the Scottish Cultural Resource Access Network (SCRAN) who, along with the ADS, will act as targets for a Historic Environs portal. It is expected to embrace a wider community eventually with expressions of interest from Europe and the USA.

The project allows the virtual searching of the holdings of the partner organisations as one. It has options to search on, in any combination, Who (creator), What (subject), When (coverage), Where (coverage) and coordinate defined geographic areas. Thus a user might cross search the ADS and RCAHMS (CANMORE) databases for references to Roman (when) forts (what) in the border area between England and Scotland (user defined coordinates).

The paper examines the technology, functionality, conformance and research potential of the portal.

## Introduction

Governments are increasingly realising the benefits of information systems. In the United Kingdom the Joint Information Systems Committee (JISC) oversees and funds the development of such systems for Higher and Further Education and facilitates resource discovery through the Distributed National Electronic Resource (DNER). In its *Guidelines for Developing an Information Strategy* (2001) the JISC notes the potential of information systems to provide '…seamless integration between local, regional, national and international information both for research and for teaching'. They envisage 'distributed computing networks providing global, institution and local access to information for everyone' with information portals and gateways as a key to providing such access.

The Archaeology Data Service (ADS) has a remit to collect, curate and provide access to digital data resulting from archaeological research in Britain and from UK funded research elsewhere. The ADS is funded in part by the JISC and forms a part of the DNER. Other funding comes from the Arts and Humanities Research Board (AHRB) and the private sector where much archaeological research takes place within the UK. Consequently data held by the ADS needs be accessible to a wider archaeological community than the strictly academic.

The wider community also holds information independent of the ADS and the academic sector with the consequence that the provision of 'seamless integration' and the provision of 'distributed computer networks' must stretch beyond the DNER. Other senses of community suggest an even wider network with a more holistic interest in the Historic Environment both using and contributing archaeological information (e.g. Baker *et al*, 1999). Similarly the Historic Environment is unlikely to respect contemporary organisational and political boundaries. For example, researching the effect of the environment on Viking settlement might necessitate using data from several disciplines collected within the contemporary boundaries of several countries.

Resource discovery has been a major concern of the Archaeology Data Service (ADS) since its inception in 1996. The ADS was heavily involved in the promotion and adoption of the Dublin Core metadata standard for categorising resources (Miller and Greenstein, 1997) and its use in facilitating interoperability between distinct resources. An early example of this was the Z39.50 enabled Gateway which performed cross searching of metadata indexes held by the various service providers (including the ADS) that comprise the Arts and Humanities Data Service (AHDS).

Research into the Historic Environment is particularly concerned with spatial and chronological attributes of data. Dublin Core provides for this with its Coverage element. Chronological descriptors such as thesaurus derived period terms or a specific date can be referenced as can locative information such as site, place, unitary authority and country. More specific geospatial information in the form of coordinates in a geographical or projected referencing system can also be recorded. The ADS provides access to some 400,000 index records about the Historic Environment, nearly all of which are spatially and temporally referenced. Co-ordinate referencing has allowed the development of various map-based search options for interfacing with ADS resources; initially relying on user input of coordinates to define a search area but more recently using click-on maps (Austin, 2000).

A recent pilot project with the Computer Laboratory of the University of Kent at Canterbury (UKC) has successfully drawn these various strands together in a portal that provides Z39.50 enabled searching of a number of geographically remote datasets. Other partners in this project include the Portable Antiquities Scheme (PAS), the Royal Commission on the Ancient and Historic Monuments of Scotland (RCAHMS) and the Scottish Cultural Resource Access Network (SCRAN) who, along with the ADS, will act as targets for a Historic Environs Portal. The RCAHMS and the PAS are outside of the academically oriented DNER. In time, the searchable network is expected to embrace an even wider community. Expressions of interest have been received from organisations in Europe and the USA.

The project builds on the pioneering work of the Aquarelle Z39.50 enabled system (Aquarelle, 1998: Michard, 1998). It allows the virtual searching of the holdings of the partner organisations as one. It has options to search on, in any combination, Who (creator), What (subject), When (coverage), Where (coverage) and co-ordinate defined geographic areas. Thus a user might cross search the ADS and RCAHMS (CANMORE) databases for references to *Roman* (when) *forts* (what) in the border area between England and Scotland (user defined coordinates). The following sections of this paper will examine the technology, functionality, standards conformance and research potential of the portal. A technical glossary is provided at the end of the paper.

## An interoperability solution

To support interoperability between diverse data sources, often with quite different internal schemas and stored data formats, a Portal needs to:

- offer a single interface to its users, irrespective of the number or type of query targets,
- translate the users' search criteria into semantically equivalent forms for each target, and
- collate and display search results when they become available.

Compliance with metadata standards is a key factor in providing interoperability between different entities as it gives well-defined semantics to create abstract levels allowing the necessary mappings.

Dublin Core (DC) is a *de facto* metadata standard. It defines a simple metadata element set consisting of 15 generic elements with well-defined semantics applicable to most, if not all, information domains. DC has been applied to several domains such as libraries, museums, geo-spatial projects, etc. The generality of elements such as creator, title, or subject illustrate its potential to be used for cross-domain searching (http://dublincore.org/ for information about DC). Some of the more specific requirements of particular domains can be met by extending DC qualifiers (DCQ) in order to provide more refined metadata elements (Dublin Core Metadata Initiative, 2000).

DC has not been free from adverse criticism having been consistently described as simplistic; however, as Weibel *et al* (1997) note it has been '*aimed at intermediate precision and high interoperability*'. More specific weaknesses have also been suggested by, for example, Godfrey Rust who see DC as fundamentally flawed as well as being simplistic especially in terms of rights management (1998). Semantically; however, it is the generality and simplicity of DC that facilitates interoperability between diverse data sets. Whilst its nature precludes the presentation of complex data structures it has seen increasing take up for resource discovery.

Z39.50 is an ISO standard defining application services and a protocol for information search and retrieval in distributed environments (NISO/ANSI/ISO, 1995). It specifies *access points* available on Z39.50 *targets* with semantics defined by information domain *profiles*. Access points specify searchable fields on virtual data sources that are managed by Z39.50 targets. Profiles are agreed pre-defined sets of rules of how to achieve interoperability by adopting well defined *attribute sets* for searching, and *abstract schemas* and *record syntaxes* for retrieving the records found in a search.

The Bath Profile is a set of rules to be adopted by the institutions providing information to obtain conformance on their Z39.50 targets to achieve interoperability (Lunau *et al*, 2001). It specifies the mandatory access points and their semantics, and the structure and syntax of returned records. Bath profile requires DC as the basic standard to provide interoperability.

Using DC, Z39.50 and the Bath profile as a basis for interoperability and the Web for accessibility, a large number of resources can be made available to users through a single portal. This Portal would be located in one institution (ADS), linking other institutions (partners) to centralise access to the partners' data sources. Users would then be able to apply high precision queries to any or all of the resources located on the connected information spaces. These access points would be accessible transparently from the Web through the Portal which connects with a target located in each entity containing access to the real fields on the databases.

This Portal aims to provide the users with a better tool to find resources either by the usual access points (e.g. Title, Subject) or by more specialised ones such as co-ordinates in a given referencing system (e.g. Referencing System, X-Coord, Y-Coord). The same results might be achieved by querying each data source using its own native query mechanism or interface, but only at a cost of considerable extra effort on the part of the user.

## Implementation

The Portal Architecture provides uniform access from the Web to the data sources, launching parallel searches over the connected targets and presenting the results as soon as they arrive. It is based on a 4-tier model, where the users gain access from any Web Browser Client. The Portal is based on a Web application implemented by Java servlets and supported by a Web server and a servlet engine.

Internally, the Portal takes the form of a Portlet developed at UKC. From the Web it appears as a Java servlet which can be accessed by HTML Forms or Applets running in a browser, or by other applications such as special-purpose clients or Web robots. This servlet accepts search parameters from the Web, converts them to suit the Z39.50 world, evaluates the queries through a special purpose query processor and launches parallel searches over all or some of the connected targets according with a mediation process, as shown in Figure 1.

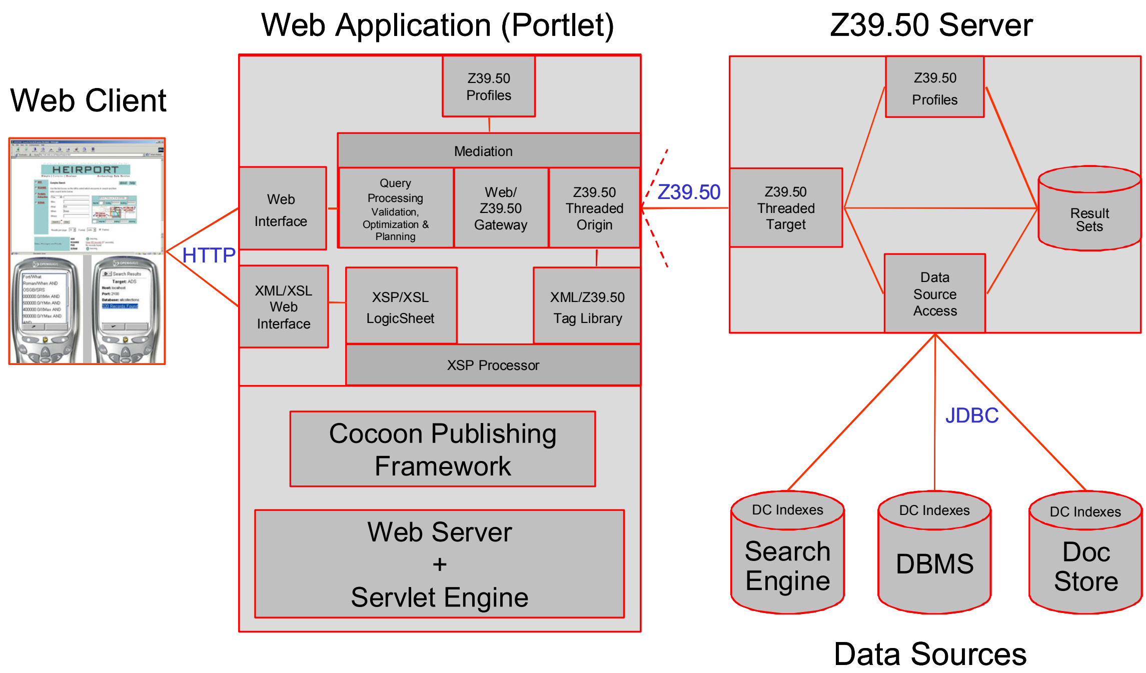


Figure 1: Portal Architecture

For Searching, HTTP requests are received by the *Portlet*, which translates them to the Z39.50 protocol using the necessary parameters (e.g. host, port, database, query-type, result set, element set, record syntax, etc). For presenting the results, HTTP requests are received by the XML/*Z39.50 TagLib*, which is supported by the *Cocoon Publishing Framework*.

Two Java packages were developed to support the Portal infrastructure, Zava and ZavaX. Zava is a Z39.50 client/server API offering basic Z39.50 features, such as Init, Search, Present and Close. Internally, Zava uses an XML parser and RDF processor to implement the profiles and to exchange the resources. ZavaX is a Web/Z39.50 API specially developed to process Z39.50 Client requests from the Web. It is supported by a Portlet for searching records according to given criteria and by a TagLib for retrieving and presenting these records. The Portlet contains a Mediator, which selects the data sources available on Targets using a query processing component, translates Web and Z39.50 transactions through a Gateway and spawns threaded searches over Z39.50 Targets using the Threaded Origin component.

The Mediator has capabilities to receive complex queries from the Web and submit them over different data sources. This mediator is an intelligent system component with crucial knowledge about the data source capabilities, which is obtained during the registration process where the data source administrator (e.g. DBA) maps the real data elements (e.g. database fields) into Z39.50 access points and record elements. Additionally, some capabilities (e.g. Spatial Referencing Systems, predefined terms belonging to thesauri) can also be registered. The query processing takes into account the mediator knowledge in order to validate, optimise and plan the queries to be submitted over the data sources.

The validation process is intended to generate a global query plan and is divided into three phases. It involves:

- checking the presence and combination of access points used in the query, which depend on the profiles implemented by the servers;
- parsing the structure of the query (e.g. boolean operators joining the sub-queries); and
- verifying the correct use of aggregated access points which only make sense if used together in order to define a given object (e.g. access points defining a spatial area within a referencing system).

The optimisation process sub-divides a given complex query into sub-queries more suitable to be dealt with by the system. It is useful, for instance, at the spatial level in order to decompose a query using the Longitude/Latitude (LL) Referencing System into sub-queries using the Ordnance Survey Great Britain (OSGB) and Ordnance Survey Northern Ireland (OSI) Referencing Systems. Therefore, data sources without LL but with OSGB and OSI capabilities will not be discarded by the mediator from a query involving the LL spatial referencing system which overlaps Great Britain and/or Ireland.

The planning process depends on the previous processes. Having a query validated and decomposed according to an optimisation scheme, an execution plan can be obtained for each data source connected to the system. Three phases are required in order to obtain the execution query plans for each data source:

- association of the query terms with *true*/*false* values (boolean operands) according to the data source capabilities for them;
- transformation of the query plan into a tree where the intermediary nodes are the boolean operators joining the sub-queries and the leaves are the boolean operands associated with the query terms; and
- evaluation of the query tree following a bottom-up strategy. Classic logic was found inadequate to process this tree, as some of the knowledge about the sub-queries is very weak. In fact, this knowledge is based on the assumption that an answer for a sub-query will not exist or may eventually exist. There is never strong evidence that a sub-query will be *true* which is required by the classic logic. In that way, the combination of boolean and set theory algebras proved to be the most adequate to process this query tree as it is based on sets that are empty (*false*) or may contain elements (*true*).

Having an execution query plan for each data source provides the system with decision support. This improves the overall system performance as it creates suitable queries for each data source avoiding over-complex queries or even submitting entire queries when it is known in advance that no answer will be returned.

The TagLib associates specific XML markup with associated logic in Java. It uses Z39.50 contextual XSP (XML Server Pages) tag values as parameters to establish a Z39.50 session and perform search and retrieval services. It is used both to receive Z39.50 parameters embedded in XML documents from the Web, and to convert records obtained from the Targets into XML. Subsequently, these XML 'records' can be transformed to any suitable format by using XSLT.

The targets expose one or more abstract databases providing access points for searching. Incoming search requests are mapped to the native queries (e.g. SQL) pointing to the DC metadata indexes on the underlying physical data sources. The result of each query generates *Result Sets*, which are managed locally by the Target and are used during the retrieval phase. Typically, a result set comprises a list of references to matching records, rather than copies of the records themselves.

Each target may have a different data sources and thus different access methods to the data sources (e.g. JDBC, ODBC, private API). Each data source has its own distinct schema. Having the same access points for the different schemas, one mapping is needed for each data source behind a target. The targets perform their respective mappings and native queries on their local databases, generate result sets and return the number of records found. As soon as the record counts arrive at the Portlet, they are made available to the user who may then choose to retrieve the records from the target.

Although Zava and ZavaX involve technologies usually not present on Z39.50 Servers, the Portlet is totally interoperable with any other Z39.50 Server. The only requirement for the Server is to implement the same profiles as the Portal. The records returned from the targets are always converted to XML for later processing by the publishing framework, even if the target does not implement the optional XML record syntax.

## HEIRPORT, an Interoperable Portal for the Historic Environment

The ADS and RCAHMS sites have Z39.50 Servers developed by UKC, whereas SCRAN and PAS have Servers developed by Systems Simulation Ltd (SSL). Having adopted the Bath profile, all Servers expose an interface consisting of well-defined DC-based access points for searching, and abstract schemas that enable retrieval according to an agreed record syntax (e.g. SUTRS, GRS1, XML).

Typical cross-domain searching uses generic access points such as *Title*, *Subject*, *Author* or even *Any* (for searching several DC elements). As the Historic Environment emphasis is on spatial and temporal searching there are also *Where* and *When* access points. This is achieved using the DC.coverage element and feeding it with temporal and spatial descriptors based on thesauri of places and time period terms. Other access points include *Who* which allows searching by DC.creator or DC.publisher, and *What* which allows searching by DC.subject element of the resource. Additional spatial access points are based on three different spatial referencing systems, OSGB (British National Grid), OSI (Irish National Grid) and LL (Latitude and Longitude geographical coordinates). All of these can be combined into complex boolean expressions to perform precise, effective and efficient searches. These access points (4Ws and Spatial) are chosen from the CIMI Profile, a special purpose application Z39.50 profile to deal with Cultural Heritage and Museum Information (Moen, 1998).

An example query based on the generic access points on this Portal might be to use a boolean option to search by What=Fort AND When=Roman at the selected targets. A more specialised query use the spatial coordinate access points based on the British or Irish national grids, or on geographical coordinates such as What=Fort, and When=Roman, Spatial_Referencing_System=OSGB, XMin=000000.0, YMin=500000.0, XMax=400000.0 and YMin=900000.0, as shown in Figure 2.


**Figure 2:** HEIRPORT Web Search Interface

Retrieval is based on result sets obtained during the search phase. During the search, a list of references to matching records is generated and stored in result sets on the target side. In response to a ‘present’ or retrieval request, the target generates and returns records in XML format. At the Portal, the Apache Cocoon Publishing Framework is then used to apply XSL Transformations to the returned XML records. In this way, the Portal can convert the XML records to a format that best suits the client application receiving the processed data. In this case, the client is normally a Web browser, so the transformation will typically generate HTML. Two examples with ‘short form’ and full records are shown in Figure 3.

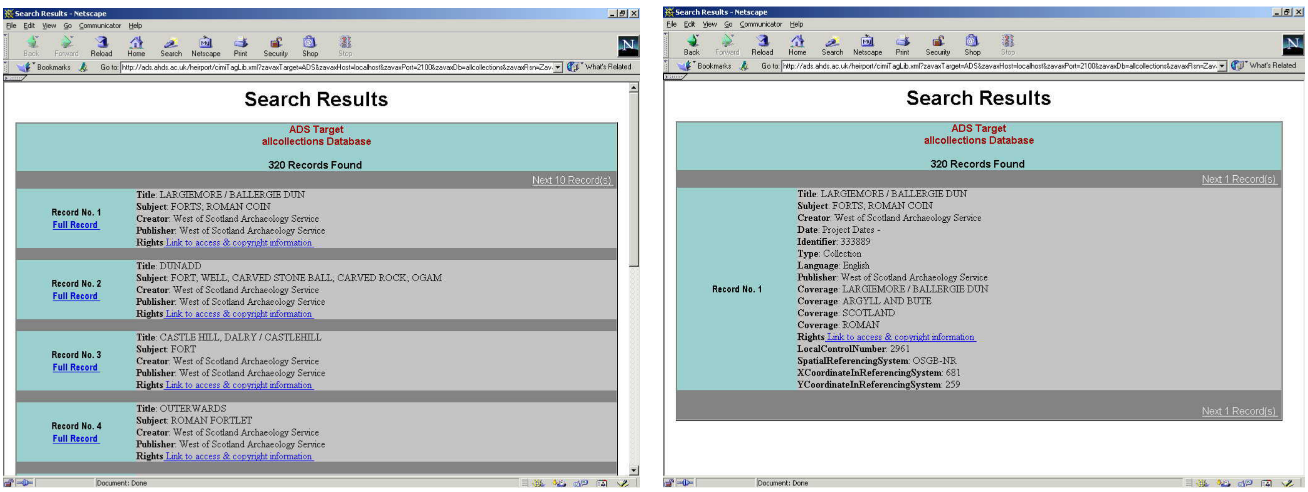


**Figure 3:** ADS Records converted to HTML in a Browser

However, as the framework has the capacity to distinguish between different browsers and other clients, it can choose to deliver different content according to the capabilities of the client. If required, it could deliver appropriate content for a small handheld computer or even for a WAP enabled mobile phone, as shown in Figure 4.


**Figure 4:** Portal Search Interface and ADS Record converted to WML in a WAP Client

The user is given the opportunity to retrieve records from each target as soon as they are available. They can therefore begin browsing through the results as soon as the first target has completed its search, and whilst other, slower, targets are still in the search phase. Appropriate timeouts are used to ensure that the user does not wait indefinitely for a target that fails to respond to the initial search request (default 10 seconds), or fails to deliver results within a reasonable time (default 3 minutes). Other safeguards are built into the targets. To prevent large-scale data theft and to limit the load imposed on the target, the maximum number of records in any query is limited and a message is returned suggesting that the user refines their query.

## Results

The main result obtained with this Portal is a solution to the interoperability problem of effective and efficient access to heterogeneous information sources.

Effective, because the access points have very precise and easily understood semantics, thus allowing users to search the Historic Environment data sets available under the targets and retrieve the records they are trying to find.

Efficient, because the searches are launched in parallel and, even if some targets are down or busy, the Portal does not wait for the last answer before presenting the results. As soon as the targets finish the local searches and send the results they are available to be retrieved from their respective sources.

The main benefit for the end-user lies in the provision of a single, simple query interface from the Web to access complex heterogeneous distributed data sources. Rather than having to locate and understand many different sources each with their own particular interface.

## Conclusions

This Portal has been developed over a period of one and a half years, involving four partners as information providers and two groups of developers. Currently, ADS, RCAHMS and PAS are already connected to the Portal and SCRAN will be connected soon.

The Portal facilitates the task of finding information as it concentrates access to all of the data sources in one place with a single interface. As Dublin Core encoded in XML is used to return the records found from each target, a common presentation of results is achieved.

In summary, HEIRPORT connects the Web with Historic Environment data sources using Z39.50 and XML, thus ensuring that such sources can play a full part in the world of XML-based interoperability.

The portal can be accessed at http://ads.ahds.ac.uk/heirport/

## Technical glossary

**American National Standards Institute (ANSI)**: American body dealing with standards http://web.ansi.org/.

**Application Program Interface (API)**: Tools and rules or protocols for building software. Blocks of code used for building programs http://webopedia.internet.com/TERM/A/API.html.

**Cocoon**: a 100% pure Java publishing framework that relies on new W3C technologies (such as DOM, XML, and XSL) to provide web content http://xml.apache.org/cocoon/.

**Database Management System (DBMS)**: A collection of programs that allows to store, modify, and extract information from a database

http://webopedia.internet.com/TERM/d/database_management_system_DBMS.html.

**Dublin Core (DC)**: An interoperable metadata standard for describing resources that enable more intelligent information discovery systems. The DC schema defines elements such as dc_creator, dc_title, dc_subject and dc_coverage. The latter is for spatial and temporal and has particular relevance to Archaeology http://dublincore.org/.

**Dublin Core Qualifiers (DCQ)**: Requirements of particular domains can be met by extending DC elements, for example, dc_coverage.period and dc_coverage.parish http://dublincore.org/documents/dcmes-qualifiers/.

**Generic Record Syntax 1 (GRS1)**: A structured information record delivery syntax for Z39.50 technology http://lcweb.loc.gov/z3950/agency/defns/oids.html#6.

**HyperText Transfer Protocol (HTTP)**: defines how messages are formatted and transmitted, and what actions Web servers and browsers should take in response to various commands. For example, when you enter a URL in your browser, this actually sends an HTTP command to the Web server directing it to fetch and transmit the requested Web page http://webopedia.internet.com/TERM/H/HTTP.html.

**Information Standards Organization (ISO)**: An international organisation for standards http://www.iso.ch/iso/en/ISOOnline.frontpage.

**Java™ Servlet**: This 'technology provides web developers with a simple, consistent mechanism for extending the functionality of a web server and for accessing existing business systems. A servlet can almost be thought of as an applet that runs on the server side -- without a face' http://java.sun.com/products/servlet/.

**Java Database Connectivity (JDBC)**: **'**An API that lets you access virtually any tabular data source from the Java™ programming language. It provides cross-DBMS connectivity to a wide range of SQL databases, and now, with the new JDBC API, it also provides access to other tabular data sources, such as spreadsheets or flat files.'

http://java.sun.com/products/jdbc/index.html.

**National Information Standards Organization NISO)**: 'develops and promotes technical standards used in a wide variety of information services. NISO is a non-profit association accredited as a standards developer by the ANSI' http://www.niso.org/.

**Open Database Connectivity (ODBC): A**n API that lets you access tabular data sources.

**Portal:** Portal or door provides search and navigation tools on the web. Today seen as providing a user or community specific context to relevant information.

**Resource Description Framework (RDF)**: A 'general framework for describing a Web site's metadata, or the information about the information on the site. It provides interoperability between applications that exchange machine-understandable information on the Web. RDF details information such as a site's sitemap, the dates of when updates were made, keywords that search engines look for and the Web page's intellectual property rights

http://www.ukoln.ac.uk/metadata/resources/rdf/.

**Structured Query Language (SQL)**: A 'standardized query language for requesting information from a database' http://webopedia.internet.com/TERM/S/SQL.html.

**Simple Unstructured Text Record Syntax (SUTRS)**: A record delivery syntax for Z39.50 technology http://www.biblio-tech.com/html/z39_50_record_syntaxes.html.

**Transmission Control Protocol/Internet Protocol (TCP/IP)**: 'Suite of communications protocols used to connect hosts on the Internet. TCP/IP uses several protocols, the two main ones being TCP and IP. TCP/IP is built into the UNIX operating system and is used by the Internet, making it the *de facto* standard for transmitting data over networks. Even network operating systems that have their own protocols, such as Netware, also support TCP/IP.'

http://webopedia.internet.com/TERM/T/TCP_IP.html.

**Wireless Application Protocol (WAP)**: 'Secure specification that allows users to access information instantly via handheld wireless devices such as mobile phones, pagers, two-way radios, smartphones and communicators' http://webopedia.internet.com/TERM/W/WAP.html.

**eXtensible Markup Language (XML)**: 'A specification developed by the W3C. XML is a pared-down version of SGML, designed especially for Web documents. It allows designers to create their own customized tags, enabling the definition, transmission, validation, and interpretation of data between applications and between organizations.'

http://www.w3.org/XML/.

**eXtensible Style Language Transformation (XSLT)**: 'A language for transforming XML documents. XSL also consists of Formatting Objects: an XML vocabulary for specifying formatting semantics. An XSL style sheet specifies the presentation of a class of XML documents by describing how an instance of the class is transformed into an XML document that uses the formatting vocabulary. Transformation is the language used in XSL style sheets to transform XML documents into other XML documents. An XSL processor reads the XML document and follows the instructions in the XSL style sheet, then it outputs a new XML document or XML-document fragment http://www.w3.org/Style/XSL/.

**ZAVA**: Java package developed by UKC; an API offering basic Z39.50 functions such as Initialisation, Search, Present and Close. Uses XML and RDF.

**ZAVAX**: Provides a mapping between XML tags and Z39.50 technology for receiving XML parameters and for the conversion of records into XML.

**Z39.50**: An ISO standard for information search and retrieval within a distributed environment. It defines application services and a protocol for communication within the distributed system (see Miller, 1999 for a useful introduction to Z39.50 technology).

## References

Aquarelle. (1998) 'Final Report', online at http://aquarelle.inria.fr/aquarelle/EN/final-report.html (downloaded 20 November 2001)

Austin, T. (2000) 'An appraisal of geospatial and chronological referencing in distributed resources within the Arts and Humanities'. *Unpub MSc dissertation, Staffordshire University.*

Baker. D., Chitty, G., Richards, J. & Robinson, D. (1999). 'HEIRNET: Mapping information resources', on-line at http://www.britarch.ac.uk/HEIRNET/rep1.html (downloaded 27 July 2001).

Dublin Core Metadata Initiative (2000), 'Dublin Core Qualifiers', online at http://dublincore.org/documents/dcmes-qualifiers/ (downloaded 27 July 2001).

JISC. (2001). 'Guidelines for developing an information strategy', online at http://www.jisc.ac.uk/pub/infstrat/ (downloaded 27 July 2001).

Lunau, C., Miller, P. & Moen, W. (2001). 'The Bath Profile: An International Z39.50 Specification for Library Applications and Resource Discovery', online at http://www.ukoln.ac.uk/interop-focus/bath/ (downloaded 27 July 2001).

Michard, A. (1998). 'Aquarelle - Sharing Cultural Heritage', *ERCIM News* 33 online edition at http://www.ercim.org/publication/Ercim_News/enw33/michard1.html (downloaded 20 November 2001)

Miller, P. and Greenstein, D. (eds). (1997). Discovering Online Resources: A Practical Implementation of Dublin Core, *UKOLN (Bath)*.

Miller, P (1999). 'Z39.50 for All', *Ariadne* 21 also online at http://www.ariadne.ac.uk/issue21/z3950/intro.html (downloaded 27 July 2001).

Moen, W. (1998). 'The CIMI Profile: Z39.50 Application Profile for Cultural Heritage Information', online at http://www.cimi.org/public_docs/HarmonizedProfile/HarmonProfile1.htm (downloaded 27 July 2001).

NISO/ANSI/ISO (1995). 'Z39.50-1995/ISO23950 Information Retrieval: Application Service Definition and Protocol Specification', online at http://lcweb.loc.gov/z3950/agency/document.html (downloaded 27 July 2001).

Rust, G. (1998). 'Metadata: The Right Approach', *D-Lib Magazine,* July/August 1998 online at http://www.dlib.org/dlib/july98/rust/07rust.html (downloaded 20 November 2001).

Weibel, S., Iannella, R. & Cathro, W. 1997. 'The 4th Dublin Core Metadata Workshop Report', *D-Lib Magazine,* June 1997 online at http://www.dlib.org/dlib/june97/metadata/06weibel.html (downloaded 20 November 2001)

## URLs of involved organisations

**Archaeology Data Service** http://ads.ahds.ac.uk/
**Arts and Humanities Data Service** http://ahds.ac.uk/
**Portable Antiquities Scheme** http://www.finds.org.uk/
**Royal Commission on the Ancient and Historic Monuments of Scotland**
http://www.rcahms.gov.uk
**Scottish Cultural Resource Access Network** http://www.scran.ac.uk/
**University of Kent at Canterbury, Comp Lab** http://www.cs.ukc.ac.uk/