

Prototyping Digital Library Technologies in *zetoc*

Ann Apps and Ross MacIntyre

MIMAS, Manchester Computing, University of Manchester,
Oxford Road, Manchester, M13 9PL, UK
ann.apps@man.ac.uk, ross.macintyre@man.ac.uk

Abstract. *zetoc* is a current awareness and document delivery service providing World Wide Web and Z39.50 access to the British Library's Electronic Table of Contents database of journal articles and conference papers, along with an email alerting service. An experimental prototype version of *zetoc* is under development, based on open standards, including Dublin Core and XML, and using the open source, leading-edge Cheshire II information retrieval technology. Enhancements investigated in this prototype include request and delivery of discovered articles, location of electronic articles using OpenURL technology, and additional current awareness functionality including the exposure of journal issue metadata according to the Open Archives Initiative protocol. These experimental developments will enhance the *zetoc* service to improve the information environment for researchers and learners.

Keywords. Electronic table of contents, current awareness, document delivery, alerting, OpenURL, Open Archives Initiative, Z39.50.

1 Introduction

The *zetoc* [1] current awareness service provides access to the British Library's [2] Electronic Table of Contents of journal articles and conference papers. It is available to researchers, teachers and learners in UK Higher and Further Education under the BL/HEFCE 'strategic alliance' [3], and to practitioners within the UK National Health Service. Access may be via the World Wide Web or the NISO Z39.50 [4],[5] standard for information retrieval which defines a protocol for two computers to communicate and share information. An experimental prototype of an enhanced version of *zetoc*, built on open standards and using open source, leading-edge software, is under development. The enhancements include ordering of copies of discovered articles, location of electronic articles and additional current awareness functionality including the exposure of journal issue metadata according to the Open Archives Initiative [6] protocol. Some of the enhancements trialled in this prototype are now implemented in the 'live' *zetoc* service. As well as being a development of a popular service, based on a significant quantity of data, the *zetoc* enhancement prototype provides a platform to explore the introduction of new technological advances to improve the information environment for researchers and learners.

2 The *zetoc* Service

The *zetoc* database contains details of articles from approximately 20,000 current journals and 16,000 conference proceedings published per year and is updated daily. With 20 million article and conference records from 1993 to date, the database covers every imaginable subject in science, technology, medicine, engineering, business, law, finance and the humanities. Copies of all the articles recorded in the database are available from the British Library's Document Supply Centre [7]. The service was developed, and is hosted, by MIMAS [8] at the University of Manchester, UK. The *zetoc* Web-Z gateway is based on that developed for the COPAC [9] research library online catalogue service which is familiar to *zetoc*'s target audience. Z39.50 compliance is provided by reworking of the COPAC application code, which utilises CrossNet's ZedKit software, developed as part of the ONE project [10], and an Open Text BRS/Search database [11]. The database is updated daily with 5000-10000 records by automatic FTP download and data conversion from the SGML format supplied by the British Library.

Searches for articles in *zetoc*, by fields such as title, author, subject and journal, may be made through the World Wide Web interface, or via Z39.50 and return details of the articles found. Currently the search results do not include abstracts for the majority of articles but there are plans to include abstracts for some articles in the future. For example, one of the results following a search in *zetoc* for articles by an author 'apps a', has a full record which includes:

```
Article Title:   Studying E-Journal User Behavior Using Log Files
Author(s):      Yu, L.; Apps, A.
Journal Title:  LIBRARY AND INFORMATION SCIENCE RESEARCH
ISSN:          0740-8188
Year:          2000
Volume:        22
Part:          3
Page(s):       311-338
Dewey Class:   020
LC Class:      Z671
BLDSC shelfmark: 5188.730000
ZETOC id:      RN083430771
```

Following a Z39.50 search, records may be retrieved as Simple Unstructured Text Record Syntax (SUTRS), both brief and full records, full records being similar to the above example, GRS-1 (Generic Record Syntax) and a simple tagged reference format. In addition *zetoc* is compliant with the Bath Profile [12], an international Z39.50 specification for library applications and resource discovery, and provides records as Dublin Core in XML according to the CIMI Document Type Definition [13].

zetoc includes a popular journal issue alerting service. Users may request email table of contents alerts to be sent to them when issues of their chosen

journals are loaded into *zetoc*. These email journal issue alerts, which are in plain text at present, list the articles and their authors within the journal issue in addition to the journal issue information. Along with each article listed is a URL which provides direct entry into the *zetoc* Web service, thus enabling the user to take advantage of the document delivery functionality of *zetoc*. Currently nearly 8000 alerts are sent out every night, and there are more than 12,500 registered users of the alerting service. Of *zetoc* Alert, Douglas Carnall in the British Medical Journal's 'Website of the week' said "The 800lb gorilla of such services in the United Kingdom is *zetoc*" [14].

An enhancement to the *zetoc* Alert service has been the introduction of alerts based on pre-defined search criteria. Users may set up an alert search request based on keywords in the article title or an author's name. These saved searches are performed on new data when it is loaded into *zetoc* each night, users being emailed with the records of articles which matched. The search-based alert is performed separately from the journal issue table of contents alert and the searches are run against all the data loaded into *zetoc*, including conference paper records.

3 The *zetoc* Enhancement Prototype

A prototype enhanced version of *zetoc* is being developed by MIMAS. It was decided to investigate a solution based on open standards and using open software. Within this version of *zetoc* the data is stored as Dublin Core [15] records, using XML syntax, generated from the SGML data supplied by the British Library. The mapping of the *zetoc* data to Dublin Core, and some of the problems associated with encoding bibliographic data using Dublin Core are described in [16].

This prototype version of *zetoc* is being used to try out enhancements before they are introduced into the service. Enhancements already added to the service, which are described in more detail below, include document delivery and subject-based alerts. In other cases, experimental enhancements employing leading-edge technology may be tested in this *zetoc* prototype but will not become part of the 'live' *zetoc* service if they are too immature. It has been found simpler to experiment with and implement these enhancements within the *zetoc* enhancement prototype, which has a flexibility provided by its use of open standard data formats and open software, rather than in the *zetoc* service which is built on proprietary data formats. It is possible that at some point in the future this *zetoc* enhancement prototype will replace the current *zetoc* service, but there is currently no timescale for this changeover.

4 The Cheshire II Information Retrieval System

The software platform used for the *zetoc* enhancement prototype is Cheshire II [17] which is a next generation online catalogue and full text information retrieval system, developed using advanced information retrieval techniques. It

is open source software, free for non-commercial uses, and was developed at the University of California-Berkeley School of Information Management and Systems, underwritten by a grant from the US Department of Education. Its continued development by the University of Berkeley and the University of Liverpool receives funding from the Joint Information Systems Committee (JISC) of the UK Higher and Further Education Funding Councils and the US National Science Foundation (NSF). Experience and requirements from the *zetoc* Cheshire prototype have been fed back into the continuing Cheshire development. Although using evolving software has caused some technical problems, the Cheshire development team has been very responsive to providing new functionality, and this relationship has proved beneficial to both projects. Examples of facilities implemented in Cheshire for *zetoc* development include sorting of result sets within the Cheshire Web interface and implementation of ‘virtual’ databases described below in section 4.3.

4.1 *zetoc* Z39.50 via Cheshire

Cheshire provides indexing and searching of XML (or SGML) data according to an XML Document Type Definition (DTD), and a Z39.50 interface. The underlying database is currently either a single file or a set of files within a directory structure, along with a set of indexes onto the data. The *zetoc* XML data is mapped to the Z39.50 Bib-1 Attribute Set [18] for indexing and searching. The Z39.50 search results formats replicate the *zetoc* service, as described above. The mapping from the *zetoc* data to the GRS-1 Tagset-G [19] elements is defined in the Cheshire configuration file for the database, and this information is used by Cheshire to return GRS-1 to a requesting client. The other Z39.50 result formats are implemented by bespoke filter programs which transform the raw XML records returned by Cheshire.

4.2 The Cheshire *zetoc* Web Interface

Cheshire also provides ‘webcheshire’ which is a basic, customisable World Wide Web interface. The web interface for the *zetoc* enhancement prototype is built on webcheshire as a bespoke program written in OmniMark (version 5.5) [20]. This *zetoc* web program provides a search interface which replicates that of the *zetoc* service, including saving session information between web page accesses, and sorting result records according to date (numeric) or title (alphabetic) using underlying Cheshire functionality. It transforms retrieved records from XML to XHTML (version 1.0) for web display. OmniMark was chosen as the programming language for this interface because it is XML (or SGML) aware according to a DTD, a knowledge which is employed for the XML translations involved. OmniMark was also chosen because of existing expertise, and the free availability of OmniMark Version 5.5 at the start of the project (but it is no longer available). Other suitable languages for the web interface implementation would have been Perl, or TCL which is the basic interface language to Cheshire.

The *zetoc* web interface provides search results in discrete ‘chunks’, currently 25 at a time, with ‘next’ and ‘previous’ navigation buttons. This is implemented by using the Cheshire capability to request a fixed number of records in the result set, beginning at a particular number within that set. The Cheshire/*zetoc* application remembers the *zetoc* identifiers of the results in the retrieved ‘chunk’, and extracts the record corresponding to a particular *zetoc* identifier when an end-user selects a ‘full record display’.

4.3 Indexing and Searching a Large Number of Records

Two problems were encountered when using Cheshire as a platform to implement *zetoc*, as a consequence of the very large and continually increasing number of records involved. These were the ordering of the returned results after a search and the size of the index files.

Sorting Result Sets. *zetoc* being a current awareness service, researchers accessing the database wish to see the most recent records first in their search results. Cheshire returns the most recently indexed results first. This is obviously not a problem for new ‘update’ data when the service is built and running, provided some thought is given to the data organisation within a directory structure. But it does mean that, if all data were in one Cheshire database, it would be necessary to load the backdata in order, with the oldest loaded first. This problem was partially resolved by the introduction of a sort capability into Cheshire which is able to return sorted results sets. By default, results returned to the user are sorted by reverse date (year of publication) order. The *zetoc* web interface provides the end-user with the option to re-sort the results by ascending date and title, and also by journal title following a journal article search. However, it becomes impractical to sort results within a very large result set, for instance of more than 500 results. Sorting larger result sets, which a dataset the size of *zetoc* could produce, would mean poor performance. Thus the problem of having to load the back data in order still remained.

Virtual Databases. Until recently all data for an application had to be indexed within a single Cheshire database. Because of concerns about this approach given that Cheshire had never been proven on this scale, a full volume test back data load was undertaken. After more than 10 million records had been indexed, performance during indexing deteriorated seriously. The problem was probably exacerbated by the fact that *zetoc* is run on the shared MIMAS service machine which is used simultaneously by other applications which process large amounts of data. By the time this experimental bulk data load was stopped, the largest of the Cheshire index files was 6 gigabytes in size. Although the operating system (Solaris 8) was able to cope with very large files, manipulation of files this size has implications for swap space and disk input/output, and hence performance.

Both of the above problems have now been addressed by the introduction of new Cheshire functionality which allows the definition of a *virtual database*.

A Cheshire virtual database contains no data itself, but is configured to search across a list of physical databases, on the same machine. A search request to a virtual database, via either Z39.50 or webcheshire, is fanned out across the underlying physical databases, and the results are assembled appropriately before display. This implementation of virtual databases supports result sets and result sorting as for a single physical database.

Using a Cheshire virtual database made it possible to load the *zetoc* data across several physical databases, one per publication year. This architecture has overcome indexing performance and index file size problems. Because the order of unsorted returned results reflects the order of the physical databases within the configuration of the virtual database, it is no longer a requirement for the back data to be loaded in publication date order, beginning with the oldest.

5 Article Delivery

5.1 Ordering from the British Library

Having discovered an article of interest in *zetoc*, a researcher will then wish to acquire the article. The *zetoc* service includes a web link to the British Library Document Supply Centre (BLDSC) which enables a user to purchase a copy of the article directly by credit card payment. *zetoc* sends the *zetoc* identifier on this link providing the BLDSC with the information it requires to locate the article within the British Library and send a copy to the customer. Within the *zetoc* enhancement prototype a demonstration link has also been included to the BLDSC Articles Direct service, for journal articles. Unlike the direct link to BLDSC, article details are filled into a web form by *zetoc*, the rest of the form being completed by the customer. In both these cases, the payment required by the British Library includes a copyright fee. Methods of delivery include mail, fax, courier, and electronic where available and agreed with the publisher concerned. This last method is currently subject to a level of paranoia amongst publishers who are refusing permission or insisting on encryption, severely constraining the usability. Exceptions are Karger, Kluwer and Blackwell Science, all of whom have given the British Library permission to supply the customer with the article in their preferred format including electronic.

5.2 Ordering via Inter-Library Loan

Within the communities which are granted access to *zetoc*, the majority of users will be entitled to order copies of articles through their own institution's Inter-Library Loan department without payment of a copyright fee, i.e. by 'library privilege', if the copy is solely for their own research or private study. In fact, 'Inter-Library Loan' is rather a misnomer, because in the case of an article a 'copy' will be supplied rather than a loan. However, the term Inter-Library Loan (ILL) is used because it is commonly understood, and it distinguishes this order method from the direct document ordering described above.

To assist researchers in making ILL requests for discovered articles, a link has been added to the *zetoc* service, following some prototyping of the facility within the *zetoc* enhancement prototype. The result of following this link depends on the institution to which the researcher belongs. Before this facility was designed, discussions were held with several ILL librarians to discover their current procedures and their opinions on what functionality *zetoc* should provide. It became apparent that there were many variations in current practices. It was also clear that some institution libraries who had developed their own forms and instructions for ILL document supply requests would want to continue to use these. On the other hand they could see the value in a researcher being able to provide the ILL department with an authoritative citation for the requested article, which also includes the British Library ‘shelf location’ information. Thus it was decided to allow institutions to choose one of two options *ILL Form* or *Local*, with a third default option for users where *zetoc* has no recorded information about their institution.

Authentication for use of *zetoc* is performed by IP address checking, and failing that by Athens [21], the UK Higher and Further Education authentication system. The Athens three-letter prefix, which is specific to each institution, is used as the institution identifier for *zetoc* ILL. ILL information from institutions and details of their library catalogue is supplied to *zetoc* support staff at MI-MAS who enter the information via a bespoke administration web form. Within the *zetoc* application, this information, including the institution identifier, is saved in an XML format ‘Institution Information’ file. When a user selects the ILL request web link, their institution identifier is determined from their login authentication and the institution information is processed as XML to provide customisation of the web pages displayed. Wording on the web pages encourages good practice, advising users to check the catalogue to determine whether their library has the article available locally, either electronically or in print, before making an unnecessary ILL request.

ILL Form Option. If an institution has requested the ‘ILL Form’ option, when a user selects the web link “Request a copy from your Institution’s Library (for research/private study)” at the foot of a *zetoc* ‘full record’ display, they are presented with a web form which includes the citation information for the discovered article. This form is in a separate web browser window to enable the user to ‘keep their place’ in *zetoc*. The user is asked to complete the form with personal details such as name, department and library card number. Selecting ‘Submit’ on this form results in a further web page containing all the captured details along with a ‘Copyright Declaration’. The user is instructed to print this form, sign it to indicate that they agree with the copyright declaration, and take it to their ILL department with the required payment. From this point onwards, the ILL request is processed according to the library’s normal procedures.

Local Option. Where an institution has requested the ‘Local’ option, selection of the ILL web link results in a page containing citation information about the

discovered article along with details of the library's ILL department. Users are instructed to use the citation information to fill out the institution's own ILL forms and follow their instructions to make the request. The 'default' option is similar, except that *zetoc* is unable to provide users with details of their ILL department.

***zetoc* Order Number.** It is possible to request only one article at once from *zetoc*, which is the article whose 'full record' displayed the selected link. This fits with ILL practice where a separate signature and payment are required for each item ordered. For all accesses to the *zetoc* ILL option, a unique *zetoc* order number is generated and appears on forms printed by the user. It was introduced following a suggestion by one institution and could have possible uses in the future. For instance if ILL requests were emailed to an institution's ILL department as well as printed by a user, the order number would allow correlation between the two modes of request.

Interoperable ILL. It is important to stress that this is the first stage only in developing a *zetoc* ILL facility. But it has set the foundations for future developments being based on: good practice; a definitive full citation for an article; existing practices; and no extra work for any party. With a view to future enhancements of this ILL facility within *zetoc*, usage will be monitored and comments from both researchers and librarians will be noted. It is likely that enhancements will be introduced incrementally as they appear to be acceptable to the community. The future vision would be to enable researchers to send their ILL document requests for articles discovered within *zetoc* directly to the British Library using a *zetoc* web form, and for the requested items to be returned using an electronic document delivery method. Recent advances such as development of a profile of the ISO ILL request format standard (ISO 10160/10161) by the Interlibrary Loan Protocol Implementers Group (IPIG) [22], which the British Library is already able to handle, and digital signatures make this vision technically nearer to realisation. But it will also be necessary to work within the existing structure of institution ILL procedures to authenticate requests and process payments from institution department budgets, which necessitates the introduction of change in a measured way.

6 Article Linking

An obvious development for a current awareness table of contents service such as *zetoc* is to provide links to the full text of an article where it is available electronically. The problem of providing such a link is two-fold if the user is to be given a link which will not be a dead end. Firstly the citation information for an article must be translated into a URL which will link to an article. Secondly this link must, if possible, be to a version of an article which is available free to the user maybe via a valid institution subscription. The latter problem is known

as that of the ‘appropriate copy’ [23]. A user would not be happy if linked to a publisher’s web site where a copy of an article was available for a substantial fee if they were entitled to read the same article through a service where they have a subscription.

6.1 OpenURL

One solution to the first of these problems is to encode a link to the full text of an article as an OpenURL [24]. OpenURL provides a syntax for transmitting citation metadata using the Web HTTP protocol and is in the process of becoming a NISO standard. The NISO committee who are developing the OpenURL standard have ‘pinned down’ the draft OpenURL as version 0.1 [25], to enable its use by early implementers, and where possible are allowing for backwards compatibility in the first version of the standard (1.0).

Within the *zetoc* enhancement prototype, OpenURLs are generated from journal article records and used for various experimental links. An OpenURL encoding using version 0.1 syntax for the journal article example shown above in section 2 would be as follows. This example shows only the ‘query’ part of the OpenURL which contains the metadata for the referent (i.e. the entity about which the OpenURL was created) and omits the resolver (BaseURL). Syntax differences between versions of OpenURL should be noted here. In a version 0.1 OpenURL the type of the referent is indicated by the label ‘genre’, as in this example. In a version 1.0 OpenURL, which may contain further entities in addition to the referent, the referent type will be defined by a ‘metadata description schema’ registered with NISO, possibly as ‘ref_valfmt=NISOArticle’ but this is not definite at the time of writing. Within this example spaces have been escape-encoded as ‘%20’ for HTTP transmission and line breaks are for clarity only.

```
?genre=article&title=LIBRARY%20AND%20INFORMATION%20SCIENCE
&atitle=Studying%20E-Journal%20User%20Behavior%20
    Using%20Log%20Files
&aulast=Yu&aunit=L
&date=2000&volume=22&issue=3
&pages=311-338&issn=0740-8188
```

6.2 Context Sensitive Linking

A solution to the ‘appropriate copy’ problem is to provide the user with a link via an OpenURL resolver which has knowledge of article subscriptions relevant to that user. Currently the best known such context sensitive reference linking service is SFX [26] from Ex Libris [27]. MIMAS are evaluating SFX as part of a separate project, ‘Implementing the DNER Technical Architecture at MIMAS’ (ITAM) [28], which includes the development of a UK academic ‘national default’ resolver. An OpenURL link to this resolver from the full record display for a journal article has been included within the *zetoc* enhancement prototype.

Following this link shows the user a range of ‘extended services’, which will include a link to the full text of the article where it is available free. Other extended services may be: a free abstract for the article; a general web search using words from the article title; a non-bibliographic function such as a library service like ‘on-line, real-time reference’ (ORR) [29]; and certain widely licensed services such as ISI ‘Web of Science’ [30] and JSTOR [31].

A link to an OpenURL resolver would be even more useful if it pointed to a resolver specific to the user’s institution. The ‘Institution Information’ XML file, described above for ILL, will be extended to include details of an institution’s OpenURL resolver, and this resolver will be used for a context sensitive link in preference to the ‘national default’ resolver.

6.3 Other OpenURL Links

OpenURL links are used to pass information internally within the *zetoc* enhancement prototype, for the ‘Articles Direct’ order and ILL order links described in section 5. Another experimental OpenURL link from a full article record in the *zetoc* enhancement prototype is to ZBLSA [32], an article discovery tool which indicates to an end-user where the full text of the article may be found, but with no guarantee of free access. An ‘OpenURL like’ link, but using proprietary labels within the URL query syntax, has been included to LitLink [33], an article linking tool from MDL Information Systems.

7 The JISC Information Environment Architecture

zetoc is part of the JISC ‘Information Environment’ [34], which provides resources for learning, teaching and research to UK Higher and Further Education, and thus must be consistent with its architecture. The Information Environment will enable article discovery through the various portals in its ‘presentation layer’, including the discipline specific Resource Discovery Network (RDN) hubs [35]. Content providers in the ‘provision layer’ are expected to disclose their metadata for searching, harvesting and by alerting. Currently the *zetoc* service provides the requisite Web and Z39.50 (Bath Profile compliant) search interfaces and an alert capability albeit in plain text. Other interfaces required of *zetoc* are OAI (Open Archives Initiative) for metadata harvesting and OpenURL for article discovery and location.

7.1 *zetoc* as an OpenURL Target

The OpenURL developments within *zetoc* described above are concerned with implementing *zetoc* as an OpenURL ‘source’, to link out from the display of the full metadata record of an article to its full text. It is also planned to implement *zetoc* as an OpenURL ‘target’ providing linking ‘in’ to the record for a specific article. It is already possible to discover an article from its metadata with a Z39.50 search to determine its *zetoc* identifier, followed by a direct web link

into *zetoc* using that identifier. Enabling *zetoc* as an OpenURL target would provide a direct web link into a particular article's record using its citation metadata. *zetoc* would then become a 'reference centre' allowing an application to provide its end-users with the ability to discover an article by a definitive citation search and then locate that article along with other relevant services.

7.2 *zetoc* as an OAI Repository

The Open Archives Initiative (OAI) has specified a Metadata Harvesting Protocol [36] which enables a data repository to expose metadata about its content in an interoperable way. The architecture of the JISC Information Environment includes the implementation of OAI harvesters which will gather metadata from the various collections within the Information Environment to provide searchable metadata for portals and hence for end-users [38]. Portals would select metadata from particular subject areas of relevance to their user community. Thus there is a requirement for collections and services within the Information Environment to make their metadata available according to the OAI protocol, including a minimum of OAI 'common metadata format', i.e. simple Dublin Core, records.

Providing an OAI interface for *zetoc* presents several problems and questions. The *zetoc* data itself is in fact metadata for the articles, and there are a very large number of records. Allowing an OAI harvester to gather all the *zetoc* records would not be sensible considering machine resources required for both the *zetoc* machine and the harvester. Harvesting *zetoc* data to provide a searchable interface would be nonsensical when the *zetoc* service itself is available for searching. In addition *zetoc* data is commercially sensitive, access restricted and owned by the British Library. There may however be some merit in making available journal issue records for harvesting. In particular, this could be useful for current awareness applications which would benefit from information about the most recent issues of journals. Thus it is intended to implement an experimental OAI interface into the *zetoc* enhancement prototype, or a data subset of it, which provides journal issue level and conference proceedings records, rather than records for articles or papers. But there is no guarantee that this will ever become a generally available service within the JISC Information Environment. Such a service would require negotiation between owners of the data and services.

A possible simple Dublin Core record provided to an OAI service for the journal issue containing the article shown above in section 2 may be as follows. An issue-level record would never contain a 'creator' but it may contain a 'contributor' if the journal issue has a named editor. The first 'identifier' in the example is the shelf location within the British Library.

```
<title>LIBRARY AND INFORMATION SCIENCE 22(3)</title>
<subject>(DDC)020</subject>
<subject>(LCSH)Z671</subject>
<date>2000</date>
<identifier>5188.730000</identifier>
<identifier>(ISSN)0740-8188</identifier>
```

<rights>All Rights Reserved
<http://zetoc.mimas.ac.uk/zetoc/terms.html></rights>

Currently *zetoc* processes article level records, holding no specific records for journal issues. In order to provide only one record for each journal issue, it will be necessary to mark the first article of an issue as such during the data load process, and include this information in the Cheshire indexes. With this data tag in place it will be possible to select issue level records when an OAI request is processed. A specific display format within the Cheshire configuration of the *zetoc* database will process the XML search result records to transform them into journal issue information.

Date Ranges. The OAI protocol allows harvesters to specify they want records ‘from’ a certain date and/or ‘until’ a certain date. Because the *zetoc* back data has been loaded recently for years gone by, using the ‘date loaded’ would not lead to sensible results. So records will be supplied according to the year of publication of the journal issue for most requests. However, it is likely that some harvesters accessing a current awareness service will want information about the latest issues of journals, which would not be readily provided using ‘year’ as the granularity. Thus if a ‘from’ date includes a month, and it is not more than two months ago *zetoc* will provide journal issues which have been added since that date. Selecting records added to the *zetoc* Cheshire database after a certain date, in response to an OAI request, is implemented easily when a Cheshire index has been created for the ‘date loaded’ field.

Acceptable Use. Implementing an OAI interface onto a very large database such as *zetoc*, which is mounted on a machine running many services, raises some concerns. With no restrictions, OAI harvesting could result in effective ‘denial of service’ attacks because of the machine resources required, and so there is a need for an ‘acceptable use’ policy. Thus there will be restrictions on how many records may be harvested at one time. When supplying only part of a result set, the OAI protocol allows for the return of a ‘**resumptionToken**’ which the harvester uses to make repeat requests. The format of this **resumptionToken** is not defined in the OAI protocol but by the source application. The **resumptionToken** from *zetoc* will include the number of the next record in the result set, along with any ‘from/until’ date information from the original request. There will be further restrictions, advertised in the information returned in response to an OAI ‘Identify’ request, on how soon a harvester can repeat the request, maybe allowing a particular harvester access only once per day. This information will also be encoded in the **resumptionToken**, to enable *zetoc* to refuse too frequent repeat requests using an HTTP ‘Retry-After’ response, or an OAI version 2 error code.

Access to the *zetoc* database through the Web and Z39.50 is restricted to particular communities by agreement with the British Library, and is authenticated either by IP address or Athens. Thus there will be tight restrictions,

requiring British Library agreement, on which services will be allowed to harvest *zetoc* data, even at the journal issue level. OAI access will be validated using IP addresses.

7.3 Current Awareness Alerting

At present the data feed for the *zetoc* Alert service is the BRS-format *zetoc* update file. In line with the other *zetoc* enhancement developments, this data feed will be changed to an XML file containing Dublin Core *zetoc* records. The search-based alerts will operate on XML records retrieved from a Cheshire database. Changing the alert data feed into an open standard format opens up the possibility of offering *zetoc* alerts in several standard formats such as XML, Dublin Core, RDF Site Summary (RSS) [37], and a tagged bibliographic format in addition to the current plain text. Providing alerts in RSS, required in the ‘Information Environment’, would enable their use for news feeds, whereas a tagged bibliographic format may be imported directly into personal bibliographic databases.

8 Conclusion

zetoc aims to provide researchers with a means to find and access published research material to aid in the furtherance of their own research, thus assisting in the advancement of knowledge. Within an internet cross-referencing paradigm of ‘discover – locate – request – deliver’ [39], the initial *zetoc* service provided discovery of research articles in a timely fashion. Enhancements to the *zetoc* service have provided ‘request and deliver’ through document supply directly from the British Library, and indirectly through traditional inter-library loan routes. Some of the experimental enhancements described in this paper indicate ways in which *zetoc* may provide ‘location’ of ‘appropriate copies’ of articles, and internet methods of ‘request and deliver’, or access, via web links.

An orthogonal purpose of *zetoc* is to provide a current awareness service through its Alert function. This service has been improved with the inclusion of search-based alerts. Further current awareness enhancements could be the provision of a choice of alert format including RSS for news feeds. The implementation of *zetoc* as an OAI repository providing journal issue records will also augment its current awareness support.

zetoc, being a popular service with a significant quantity of data, has provided a platform to prototype new technologies and possible additions to the service. One aim of the ‘*zetoc* Enhancement Project’ was to develop a solution based on open standards and using leading-edge, open source technology. This has been successfully achieved within a prototype environment using a Cheshire II software platform to index *zetoc* Dublin Core records encoded in XML. A spin-off has been improvements to Cheshire following *zetoc* feedback. Other experimental technologies such as OAI and OpenURL will enable *zetoc* to be integrated into the JISC ‘Information Environment’, thus providing a valuable service to the stakeholders within that environment.

Acknowledgements. The authors wish to acknowledge the contribution to the development of *zetoc* by their colleagues at the British Library, including Stephen Andrews and Andrew Braid, at MIMAS, Ashley Sanders, Jane Stevenson, Andrew Weeks and Vicky Wiseman, and the Cheshire development team, Ray Larson at the University of California–Berkeley and Paul Watry and Robert Sanderson at the University of Liverpool. The initial development of the *zetoc* service was funded by the British Library who own and supply the Electronic Table of Contents data. The ‘*zetoc* Enhancement Project’ is funded by the British Library and by the Joint Information Systems Committee (JISC) [40] for the UK Higher and Further Education Funding Councils, as part of the ‘Join-Up’ programme [41] within the Distributed National Electronic Resource (DNER) development programme [42].

References

1. *zetoc*, Electronic Table of Contents from the British Library. <http://zetoc.mimas.ac.uk>
2. The British Library. <http://www.bl.uk>
3. Strategic alliance emphasises British Library’s central role in support of higher education. Press Release, 19 March 2002. <http://www.bl.uk/cgi-bin/press.cgi?story=1231>
4. Z39.50, the North American National Information Standards Organisation (NISO) standard for information retrieval. <http://www.niso.org/z3950.html>
5. Miller, P.: Z39.50 for All. *Ariadne* **21** (1999). <http://www.ariadne.ac.uk/issue21/z3950>
6. Open Archives Initiative (OAI). <http://www.openarchives.org/>
7. British Library Document Supply Centre (BLDSC). <http://www.bl.uk/services/document/dsc.html>
8. MIMAS, a UK Higher and Further Education data centre. <http://www.mimas.ac.uk>
9. The COPAC research library online catalogue service. <http://copac.ac.uk>
10. CrossNet ZedKit software. <http://www.crxnet.com>
11. Open Text BRS/Search. <http://www.opentext.com/dataware/>
12. The Z39.50 Bath Profile. <http://www.nlc-bnc.ca/bath/bp-current.htm>
13. The Consortium for the Computer Interchange of Museum Information (CIMI) Dublin Core Document Type Definition. <http://www.nlc-bnc.ca/bath/bp-app-d.htm>
14. Carnall, D.: Website of the week: Email alerting services. *British Medical Journal* **324** (2002) 56
15. The Dublin Core Metadata Initiative. <http://www.dublincore.org>
16. Apps, A., MacIntyre, R.: *zetoc*: a Dublin Core Based Current Awareness Service. *Journal of Digital Information* **2**(2) (2002). <http://jodi.ecs.soton.ac.uk/Articles/v02/i02/Apps/>
17. The Cheshire II Information Retrieval System. <http://cheshire.lib.berkeley.edu>
18. The Z39.50 Bib-1 Attribute Set. <http://lcweb.loc.gov/z3950/agency/defns/bib1.html>
19. The Z39.50 Generic Record Syntax (GRS-1) Tagsets. <http://lcweb.loc.gov/z3950/agency/defns/tag-gm.html>

20. OmniMark Technologies. <http://www.omnimark.com>
21. Athens Access Management System. <http://www.athens.ac.uk>
22. Interlibrary Loan Protocol Implementers Group (IPIG) Profile for the ISO ILL Protocol. <http://www.nlc-bnc.ca/iso/ill/ipigprfl.htm>
23. Caplan, P., Arms, W.Y.: Reference Linking for Journal Articles. *D-Lib Magazine* 5(7/8) (1999). doi://10.1045/july99-caplan
24. OpenURL, NISO Committee AX. <http://library.caltech.edu/openurl/>
25. OpenURL Syntax Description (v0.1). <http://www.sfxit.com/OpenURL/openurl.html>
26. Van de Sompel, H., Beit-Arie, O.: Open Linking in the Scholarly Information Environment Using the OpenURL Framework. *D-Lib Magazine* 7(3) (2001). doi://10.1045/march-vandesompel
27. Ex Libris, SFX Context Sensitive Reference Linking. <http://www.sfxit.com>
28. 'Implementing the DNER Technical Architecture at MIMAS' (ITAM) project. <http://epub.mimas.ac.uk/itam.html>
29. Moyo, L.M.: Reference anytime anywhere: towards virtual reference services at Penn State. *The Electronic Library* 20(1) (2002) 22-28.
30. ISI Web of Science Service for UK Education. <http://wos.mimas.ac.uk>
31. JSTOR, the Scholarly Journal Archive (UK). <http://www.jstor.ac.uk>
32. ZBLSA - Z39.50 Broker to Locate Serials and Articles. <http://edina.ac.uk/projects/joinup/zblsa/>
33. LitLink, MDL Information Systems. <http://www.litlink.com>
34. Powell, A., Lyon, L.: The JISC Information Environment and Web Services. *Ariadne* 31 (2002). <http://www.ariadne.ac.uk/issue31/information-environments/>
35. Resource Discovery Network (RDN). <http://www.rdn.ac.uk>
36. Warner, S.: Exposing and Harvesting Metadata Using the OAI Metadata Harvesting Protocol: A Tutorial. *High Energy Physics Libraries Webzine* 4 (2001). <http://library.cern.ch/HEPLW/4/papers/3/>
37. Powell, A.: RSS FAQ, JISC Information Environment Architecture. <http://www.ukoln.ac.uk/distributed-systems/dner/arch/faq/rss/>
38. Cliff, P.: Building ResourceFinder. *Ariadne* 30 (2001). <http://www.ariadne.ac.uk/issue30/rdn-oai/>
39. MODELS - Moving to Distributed Environments for Library Services. <http://www.ukoln.ac.uk/dlis/models/>
40. The Joint Information Systems Committee (JISC). <http://www.jisc.ac.uk>
41. The Join-Up programme. <http://edina.ed.ac.uk/projects/joinup/>
42. The UK Distributed National Electronic Resource (DNER). <http://www.jisc.ac.uk/dner/>