

## Working with the British Library - the '*zetoc*' experience

Ross MacIntyre and Ann Apps  
MIMAS, Manchester Computing, University of Manchester  
Oxford Road, Manchester, M13 9PL, UK  
Email: ross.macintyre@man.ac.uk, ann.apps@man.ac.uk

### Abstract

A market research report commissioned by the British Library was used as the basis of a detailed list of requirements for the provision of digital services to the UK Higher and Further Education sector. The services stemming from these requirements centred, in the first instance, on the British Library's Electronic Table of Contents data (ETOC) that lists the titles of nearly 15 million journal articles and conference papers. A key approach to the development of these services will be the conformance to accepted standards and open systems, thus enabling greater potential for interoperability with developments elsewhere. As a first step towards this the British Library contracted Manchester Computing to develop and mount a Z39.50-compliant version of the ETOC database, a service now live entitled "*zetoc*". The service is free to UK institutions supported by the Joint Information Systems Committee (JISC) of the UK Higher Education Funding Councils. In addition to this operational service, which is based on existing and well-established proprietary database management applications, a separate pilot version is being developed using a subset of the data in XML format which will be conformant with the Dublin Core and exploiting 'Cheshire II' software.

As well as a number of Z39.50 interface developments, a current awareness alerting service based on the ETOC data has been implemented. The ultimate aim is to develop the alerting service to enable it to link seamlessly with document ordering systems and integrate with other current awareness mechanisms. The document ordering interface will present the user with the option to select a source of supply. If the item is contained within an electronic journal to which the user's institution subscribes, the intention is to automatically link the user to the electronically-stored journal. If the item is not held 'locally', the user will be presented with an option to either request the document directly from the British Library or via their local library. This functionality will be achieved via integration with other systems, including 'brokering' services, developed within the UK education community as part of the JISC's accurately named 'Join-Up' programme.

This paper describes the development process, the systems themselves and the support services created to provide academic and research communities with a valuable resource both in terms of discovery/location and content provision. The British Library is working in partnership with Manchester Computing at the University of Manchester, with the Universities of Liverpool and California-Berkeley as associate partners.

## 1. Background

The British Library (BL) is the UK's national library and has had a close relationship with the UK Higher Education (HE) community for a long time. They are the prime source for Document Supply and Inter-Library Loan (ILL) requests and have been active in many other areas, such as the MODELS [1] series of workshops.

More recently, the British Library has placed an increasing strategic emphasis on collaboration to deliver its objectives. In particular, the BL has been seeking closer working ties with UK HE. In April 2000 a joint BL/HE Task Force was announced [2] with the aim of identifying "specific initiatives for mutual benefit"

Prior to the formation of this Task Force, the BL had already approached MIMAS [3], located at the University of Manchester, to identify and discuss possible collaborations. MIMAS is the UK's largest national academic data centre and receives funding from the Joint Information Systems Committee, JISC [4], of the Higher and Further Education Funding Councils to provide services to the academic community in the UK and beyond.

The BL had in mind to offer to licence their *inside* product free of charge to the UK's HE, FE and Research communities. The product was already in use at a number of institutions and these were to be refunded their licence fee.

At the heart of the *inside* product is the 'ETOC' database, containing approximately 16 million records corresponding to journal articles or conference papers. The data goes back to 1993 and covers 20,000 journal titles and 16,000 conferences per year. The entries are mostly keyed in, typically within 72 hours of publication and the database is updated daily with around 10,000 new entries. The database itself does not contain the full-text of the articles/papers, nor does it contain many abstracts. However, the presence of a record indicates that the BL have a copy for supply. The journal titles represent those most often requested and as UK HE generate approximately 50% of these requests, reflect what that community needs.

The BL/MIMAS discussions led to MIMAS signing a formal contract with BL covering:

- assisting with a Document Supply requirements analysis
- developing a service using the ETOC data by the 2000/2001 academic year
- prototyping a service based on open standards

The contract period was April 2000 through March 2001 and the BL were to fund all the above, though JISC were to be approached about on-going funding of the service, which they subsequently agreed to do.

The requirements analysis was to be derived via a market research exercise involving a specialist firm, Orbitel Marketing, during April and May 2000. The results have subsequently been made available to the community and have been adopted by the 'Join-Up' Programme [5], which is mentioned later. The report itself is not described further in this paper.

The ETOC-based service was to be Z39.50-compliant (which *inside* was not) and include a Web interface, for those users without Z-client software. The service was soon referred to as 'ZETOC' and later trademarked by the BL as *zetoc*.

The prototype was to be based on XML and open source application code for indexing, access and Z39.50 compliance. Relevant standards were to be adhered to wherever possible.

## 2 Development Approach

Counting back from September 2000 it was apparent that the application and service components needed to be developed 'quickly'.

It was decided to reuse existing software and application code where appropriate and to involve staff already employed within the Manchester Computing department. Similarly, there was a need to minimise any training and support documentation requirements that would be required for the new service. This led to the service being based on a user interface from a well-used service, COPAC [6], the UK's research libraries online catalogue, that was familiar to the target audience. Also, Z39.50 compliance could be offered by reworking the COPAC application code, which utilised CrossNet's ZedKit software, developed as part of the ONE project [7] and a BRS/Search [8] database.

However, there was the opportunity to include relevant feedback MIMAS had received regarding other services, for example that people do not like having limits set on the number of 'hit records' they can have when searching.

The prototype, consisting of a subset of the full ETOC records, was to utilise Cheshire II software [9], which was developed at the University of California-Berkeley School of Information Management and Systems, underwritten by a grant from the Department of Education. Its continued development at the two Universities of Berkeley and Liverpool receives funding from JISC and the US NSF (National Science Foundation). This was chosen because it offered a more flexible and open development path than the BRS version.

## 3 *zetoc* Service Development

The development began with a data analysis exercise. This covered both the data to be held in the database itself and the daily update files.

The BL actually also used BRS/Search for the *inside* service and this simplified the mapping, though certain changes were to be made, principally stripping out BL's internal processing data. The BL extracted the data into files in 6-month 'chunks' and MIMAS converted and loaded the data into a newly created BRS database. The data design and database placement took most of July and the subsequent bulk loading process took all of August to complete.

The BL supply update files in SGML format. MIMAS developed an SGML DTD for the data and developed code to convert the data into the required BRS load file

format. This daily update is now a totally automated process, including the ftp download from the BL.

The Web application was developed essentially as a search facility, offering three search options:

1. General. This option allows you to search the whole database of journal articles and conference papers by article/paper title, author(s), ISBN/ISSN and year of publication. For example you might wish to see what articles/papers are available by a particular author, or you might search for the details of a specific document using the author and title.
2. Journal. This option allows you to search for journal articles by article title, author(s), journal title, volume, issue, part, start page of article and year of publication.
3. Conference. This option limits your search to conference proceedings only. You can search by paper title, author(s), keywords and conference details (conference name, sponsor, venue and date held).

As an example, a search in *zetoc* for articles by an author “apps a”, results in a list of brief search results including:

Dublin Core Metadata for Electronic Journals / Apps, A.; MacIntyre, R.  
LECTURE NOTES IN COMPUTER SCIENCE - 2000; ISSU 1923; Pages:  
93-102

Following a link beside this brief record displays a more detailed record for the article:

Article Title: Dublin Core Metadata for Electronic Journals  
Author(s): Apps, A; MacIntyre, R  
Journal Title: LECTURE NOTES IN COMPUTER SCIENCE  
ISSN: 0302-9743  
Volume: 1923  
Year: 2000  
Jnl Issue Title: Research and Advanced Technology for Digital Libraries  
Page(s): 93-102  
Editor(s): Borbinha, J; Baker, T  
Publisher: Germany : Springer-Verlag  
Language: English  
Dewey Class: 004  
LC Class: TP372.5  
BLDSC shelfmark: 5180.185000  
ZETOC ID: RN085008791

End-users of the Web service may request discovered records to be emailed to them.

In addition to the Web search interface, *zetoc* also has a Z39.50 interface. It allows for searching via the Z39.50 Bib-1 Attribute Set [10].

It will return information as SUTRS (Simple Unstructured Text Record Syntax), GRS-1 (Generic Record Syntax) [11] and a simple tagged reference format. In order to be “Bath Profile compliant”[12], referring to the specification for library applications and resource discovery, *zetoc* also has the option to return simple Dublin Core records

encoded according to the CIMI Dublin Core DTD [13] using an XML syntax. The mapping of most of the fields in a *zetoc* record is obvious, but there are issues with returning bibliographic citation information in this format [14].

The SUTRS format is similar to that displayed as the result of a search using the Web interface, but as plain text without the HTML tags. The simple tagged format returns fields of the record preceded by a token, e.g. TI: precedes a title, again in plain text. This format may be used for importing citations into a personal bibliographic database.

Access to the database is via <http://zetoc.mimas.ac.uk> and controlled by IP/domain address filtering and ATHENS authentication [15], a standard authentication system for UK resources.

The *zetoc* service was launched on schedule on 26th September 2000 at the British Library, St Pancras, London. As well as being available to British and Irish higher and further education institutions and Research Councils, access has recently been extended to the NHS via the NeLH (National Electronic Library for Health). Additionally the use of the Z39.50 target is currently being trialled by the CIC (Committee on Institutional Cooperation) consortium in the US.

#### 4 *zetoc* Alerting Service

To supplement the basic search and retrieve functionality of the service, a current awareness alerting service based on the table of contents data was also developed. The aim was to 'fill the gap' left by the demise of the popular 'Autojournals' service offered by BIDS, another of the UK's national academic datacentres, until July 2000. The application had to work with both the operational and prototype versions of the database so as not to prejudice any future decision concerning the service's technical architecture.

The service allows the user to create one or more named lists of journal titles of interest. There are three ways to choose journals:

1. Select journal names beginning with a letter - if you are looking for a specific journal, select the first letter of the title to view an alphabetical list of journals beginning with that letter.
2. Select journal names containing a string - if you know part of the journal name, type it in the box and select the Search button. Note that this is **not** a keyword search.
3. Select journals by subject category - the journals in *zetoc* have been grouped into subjects according to the Dewey Decimal subject classification. Select one of the subjects and you will be given a list of journal titles in that category.

They are then sent the Table of Contents of newly loaded issues via email. This process is driven by the daily update of the main database.

An example extract of an Alert email when first released:

Subject: ZETOC Alert: NEW SCIENTIST

ZETOC Alert results for list MYSCIENCE

NEW SCIENTIST  
ISSU 2255; 15 September 2000  
ISSN 0262-4079

24-29

Titanic You may think today's stars are awesome, but compared with the first  
suns they're mere whippersnappers  
Chown, M.

30-33

Hairy space probes What's purple, furry and sensitive all over?  
Brooks, M.

The first version of *zetoc* Alert was released in October 2000. Subsequently each  
article was linked to its corresponding *zetoc* record via a URL of the form:

<http://zetoc.mimas.ac.uk/zetoc/wzgw?terms= RN085008791&field=zid>

The purpose being to allow the user to move directly to the record and from there take  
advantage of full system functionality without having to duplicate features, full article  
access being the obvious future goal.

## 5 Support Facilities

As well as the service application development, documentation was produced to assist  
the users and library support staff. As the application was specifically designed to be  
simple and familiar, there has been little need to produce more than an on-line user  
guide and workbook.

The BL ran a series of training/familiarisation sessions throughout the UK during  
2000 and 2001. The services have received promotion via the well-established 'JIBS'  
User Group [16], which keeps institutions up-to-date with new JISC services and  
where they may be useful. JIBS are also a key route for feedback from the community  
to service and content providers alike.

A range of statistical measures have been generated as part of the service and are  
accessible via the Web site [17]. For the *zetoc* service there are monthly usage figures  
broken down by domain and Web versus Z39.30 access. For "ac.uk" domains these  
are further broken down by individual institution. The figures are available as tables,  
graphs and as a 'csv' format file for downloading.

Alert statistics are updated each day and show number of: users (6644), lists (9668),  
journal titles selected (107435) and emails sent (3189). (Figures in brackets are as on  
9<sup>th</sup> August 2001). The journal title count shows that there are an average of over 11  
journals on each list and each user has on average 1.5 lists. There is also a monthly

'Top 20' list of journals requested. As of 9<sup>th</sup> August the top 5 were: Nature, Science, Lancet, British Medical Journal and Journal of Academic Librarianship.

## 6 *zetoc* Prototype

MIMAS and the British Library are now working on an enhanced version of *zetoc*, which is currently a prototype. 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 records, using an XML syntax, generated by bespoke programs from the supplied British Library SGML. This XML is indexed using the Cheshire open source software. Cheshire II is a next generation online catalogue and full text information retrieval system. It was developed using advanced information retrieval techniques and provides customisable Web and Z39.50 interfaces.

It is the intention to use this prototype version of *zetoc* to trial enhancements to the service, such as the facility to order, or link to, the full text of discovered articles, and subject-based alert requests. Within an Internet cross-referencing paradigm of 'discover – locate – request – deliver' the present *zetoc* current awareness service provides discovery of research articles in a timely fashion. Early enhancements to *zetoc* will provide 'request and deliver' through document supply from the British Library.

Future enhancements may include 'locate' of the appropriate copy, through an initiative such as SFX Content Sensitive Reference Linking [18] or other resolution services, and 'request' and 'deliver' via internet linking mechanisms, including Digital Object Identifiers [19] and CrossRef [20], to freely available articles or those covered by an institutional subscription. It will be simpler to implement these, and future, enhancements with the data held in open standard formats such as Dublin Core and XML.

## 7 Future Development

Four areas of service development have been highlighted below.

### 1) Document Ordering

The service will be enhanced during September 2001 to include the facility to request the article/paper from the British Library's Document Supply Centre (BLDSC). Initially this would require the user to pay a copyright fee, as this was technically the simplest facility to enable. However, in early 2002 Inter-Library Loan requests will be supported. The emphasis will be on ISO ILL formatted requests (ISO 10160/10161), which can already be handled by the BL.

### 2) Subject/Author Alerting

The Alert service will allow users to enter key words/phrases and/or author names. The daily update file will be searched for occurrences and hit records be sent to the

user via email. An example extract is shown below, where “fish” has been entered as a title keyword.

Subject: zetoc Alert: (fish)[ti]  
zetoc Alert results for list fishtest belonging to manfish

CANCER CAUSES AND CONTROL VOL 12; PART 4 pp. 375-382  
A pooled analysis of case-control studies of thyroid cancer. VI. Fish and shellfish consumption  
Bosetti, C.; Kolonel, L.; Negri, E.; Ron, E.; Franceschi, S.; Maso, L. D.; Galanti, M. R.; Mark, S. D.; Preston-Martin, S.; McTiernan, A.  
<http://zetoc.mimas.ac.uk/zetoc/wzgw?terms=RN097856709&field=zid>

### 3) Bibliographic Data

Within the *zetoc* service the user will be able to request records be emailed in short tagged format, suitable for loading into a personal bibliographic database. For example:

AU: Noga, E. J.  
AU: Fan, Z.  
AU: Silphaduang, U.  
TI: Histone-like proteins from fish are lethal to the parasitic dinoflagellate *Amyloodinium ocellatum*  
JT: PARASITOLOGY -CAMBRIDGE-  
IS: 0031-1820  
PD: 2001  
IU: VOL 123; PART 1  
PG: 57-66  
FQ: Bi-monthly: 5-8 issues per year  
PB: CAMBRIDGE UNIVERSITY PRESS  
PP: Great Britain  
LA: English  
DC: 616.96  
LC: QL  
SM: 6406.000000  
ID: RN098262714

### 4) The ‘Join-Up’ Programme

Subsequent to the May 2000 meeting of the BL/HE Taskforce, the BL was invited to discuss the perceived overlap between some of the projects submitted for JISC funding and the BL’s own development plans. The outcome of the meeting was that the respondents were asked to try and present their projects within a more coherent framework, as each specifically addressed one or more of the ‘discover – locate – request – deliver’ functions. The projects, XGRAIN (a Z39.50 end-user interface to A&I databases), ZBLSA (a locator mechanism based on serials information) and Docusend (a locate and document delivery system based on Fretwell-Downing’s VDX system) together with *zetoc* have since been drawn together formally into “Join-Up”.

The *zetoc* service can support “Join-Up” through work that is already underway or planned:

- the development of a Z39.50 target on ETOC thus enabling it to interface with XGRAIN
- the use of ZBLSA as a serials-based locator for document ordering purposes
- implementation of the ISO/ILL protocol as a way of accepting requests from Docusend
- alerting as an additional facility within the overall system.

Any duplication with Docusend with regard to document ordering functionality can be reconciled in that the prototype implementation with the Cheshire II software is an ‘open systems’ alternative to the proprietary system of Docusend.

The projects have also agreed to work together to identify common deliverables and goals, the area of evaluation being an obvious candidate.

## 8 Conclusions and Observations

Overall, the collaboration has been very successful in that a new service was developed and made available in very short timescales and for a very small cost to the community. The following comments apply to various aspects of the service development.

### Content is King

The area of content provision is clearly one of the major strengths that the British Library can bring to benefit the community via initiatives such as the *zetoc* service. In turn, the BL have stated that they view *zetoc* as a key part of the British Library's strategy to open up access to its collections (in this instance its Electronic Table of Contents) through networked services, especially to UK HE. So simplistically it seems like a ‘win/win’ situation.

### Application Development

The reuse of existing application code supplemented by the use of open standards and a modular design has enabled very rapid and efficient service development. The non-duplication of functionality has been possible by enabling the applications to communicate, c.f. interoperate. Looking forward, the number of opportunities enabled by such an approach are also significant.

### KISS

The use of a simple and familiar interface reduced drastically the need for user support and training. This is borne out by the end user enquiries received by the MIMAS help desk, which tend to be about people wanting access to either the service or to the full text of the articles, rather than problems understanding or using the system’s features. The old adage of KISS – “Keep It Simple, Stupid!” remains valid.

### Feedback

It is important to make it easy for people to comment constructively, provided that feedback is then acted upon. Always provide a mechanism to allow people to supply unstructured comments and do not assume all users will have access to an email client

at their workstation. It has been noticeable that people will use these routes to say “thank you”, i.e. provide positive feedback. Relevant user groups, such as JIBS, should be used to provide expertise in explaining, classifying and prioritising the requirements.

### Service Perspectives

It is interesting to note a difference in perception of the service from the perspectives of the BL and the end-user. The BL regard the database as a list of (a very small part of) its holdings and thus view *zetoc* as a document ordering facility. Whereas the end user seems to regard the service primarily as a current awareness tool, allowing them to be notified of new articles and papers and to find others from partial information.

### 9 Acknowledgements

The authors wish to acknowledge the contribution to the development of *zetoc* by their colleagues at the British Library, Stephen Andrews and Andrew Braid, and at MIMAS, Alison Murphy, Ashley Sanders, Andrew Weeks and Vicky Wiseman. 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) for UK higher and further education, as part of the Join-Up programme within the Distributed National Electronic Resource (DNER) development programme [21].

### 10 References

- [1] MODELS – Moving to Distributed Environments for Library Services  
<http://www.ukoln.ac.uk/dlis/models/>
- [2] BL/HE Task Force [http://www.jisc.ac.uk/press\\_releases/000524\\_blhe.html](http://www.jisc.ac.uk/press_releases/000524_blhe.html)
- [3] MIMAS – Manchester InforMation and Associated Services  
<http://www.mimas.ac.uk>
- [4] JISC – Joint Information Systems Committee <http://www.jisc.ac.uk>
- [5] Join-Up Programme <http://edina.ac.uk/projects/joinup/index.shtml>
- [6] COPAC – CURL OPAC <http://www.copac.ac.uk>
- [7] CrossNet’s ZedKit software <http://www.crxnet.com>
- [8] BRS/Search <http://www.opentext.com/dataware/>
- [9] The Cheshire II Information Retrieval System. <http://cheshire.lib.berkeley.edu>
- [10] The Z39.50 Bib-1 Attribute Set.  
<http://lcweb.loc.gov/z3950/agency/defns/bib1.html>
- [11] The Z39.50 Generic Record Syntax (GRS-1) Tagsets.  
<http://lcweb.loc.gov/z3950/agency/defns/tag-gm.html>

- [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] A. Apps and R. MacIntyre. (2001). *zetoc*: a Dublin Core Based Current Awareness Service. Proceedings of the International Conference on Dublin Core and Metadata Applications 2001 (DC2001 - <http://www.nii.ac.jp/dc2001/>), National Institute of Informatics, Tokyo, Japan, 22-26 October 2001 (accepted for publication).
- [15] ATHENS <http://www.athens.ac.uk>
- [16] JIBS – JISC (assisted) Bibliographic dataseviceS' <http://hosted.ukoln.ac.uk/jibs/>
- [17] *zetoc* Usage Statistics <http://zetoc.mimas.ac.uk/stats/>
- [18] SFX <http://www.sfxit.com>
- [19] H. Atkins, C. Lyons, H. Ratner, C. Risher, C. Shillum, D Sidman and A. Stevens. Reference Linking with DOIs. D-Lib Magazine 6(2), February 2000.  
doi://10.1045/february2000-risher
- [20] CrossRef. <http://www.crossref.org>
- [21] DNER – Distributed National Electronic Resource <http://www.jisc.ac.uk/dner/>