

zetoc: a Dublin Core Based Current Awareness Service

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

Abstract

zetoc is a current awareness service for UK Higher and Further Education providing Z39.50 access to the British Library's Electronic Table of Contents database of journal articles and conference papers. The **zetoc** database, updated daily, may also be searched via a World Wide Web interface. An alerting service provides tables of contents by email for new journal issues when they are loaded. The current version of **zetoc** is Z39.50 Bath Profile compliant and can provide Dublin Core records encoded in XML in answer to Z39.50 search requests. An enhanced version of **zetoc**, currently a prototype under development, will hold the data within an XML repository, using Dublin Core as the basis of the metadata schema. This paper describes the encoding of bibliographic records for journal articles and conference papers in Dublin Core, and the interoperability between Dublin Core and other bibliographic standards.

Keywords: Dublin Core, bibliographic citations, metadata standards, current awareness.

1 Introduction

The **zetoc** [1] current awareness service provides access to the British Library's [2] Electronic Table of Contents, primarily for researchers, teachers and learners in UK Higher and Further Education. Access may be via the World Wide Web, or via the NISO Z39.50 [3,4] standard for information retrieval which defines a protocol for two computers to communicate and share information. It is compliant with the Bath Profile [5], an international Z39.50 specification for library applications and resource discovery. A prototype of an enhanced version of **zetoc**, with the data encoded using the Dublin Core Metadata Element Set [6,7] within an XML [8] syntax, is now under development. The enhancements include document delivery to provide ordering of the discovered documents from the British Library. The enhanced version of **zetoc** is based on open standards and uses open software. As well as being a development of a popular service, based on a

significant quantity of data, it provides a platform to explore the use of Dublin Core for bibliographic records and to investigate the interoperability between several standard metadata systems.

2 The zetoc Service

The **zetoc** database contains details of approximately 20,000 current journals and 16,000 conference proceedings published per year and is updated daily. With almost 15 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. The service was developed, and is hosted, by MIMAS [9] at the University of Manchester, UK. The **zetoc** Web-Z gateway is based on that developed for the COPAC [10] research library online catalogue service. The **zetoc** data is held in a Dataware BRS/Search [11] database. BRS/Search is a longstanding and reliable, but proprietary, information retrieval system. The database is updated daily with 5000-10000 records by automatic FTP download, data conversion and loading each night. Data is supplied by the British Library in SGML format which is translated by conversion programs, written in-house, into the required data load format for BRS.

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; VOL 1923; Page(s) : 93-102 [Research and
Advanced Technology for Digital Libraries]

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 may request discovered records to be emailed to them. Similar records can be discovered via the Z39.50 Simple Unstructured Text Record Syntax (SUTRS) interface.

zetoc includes a 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. Currently about 3500 alerts are sent out every night, and there are more than 6000 registered users of the alerting service.

The only use of Dublin Core in the current service is the Z39.50 XML option which returns simple Dublin Core records encoded according to the CIMI Dublin Core Document Type Definition [12] as the result of a search on the **zetoc** database, as required for Z39.50 Bath Profile compliancy. 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 which are discussed in more detail below.

3 An Enhanced Version of **zetoc**

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 [13] open source software developed at the University of Berkeley. 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 World Wide 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, possibly through an initiative such as SFX Content Sensitive Reference Linking [14] or other resolution services, and 'request' and 'deliver' via internet linking mechanisms, including Digital Object Identifiers [15] and CrossRef [16], to freely available articles or those covered by an institutional subscription. It should be simpler to implement these, and future, enhancements with the data held in open standard formats such as Dublin Core and XML.

4 Data Mapping to Dublin Core

Most of the fields of the **zetoc** records map obviously to Dublin Core elements. They include article title, authors, subject codings using Library of Congress and Dewey classifications and the publication year ('issued' date). For a conference paper there are some additional conference subject keywords. Some fields contain identification specific to the British Library such as the shelf location. Where appropriate, Dublin Core qualifiers [17] are used, but some additional **zetoc**-specific qualifiers are employed. In some cases, sub-elements within a **zetoc** namespace are used, for instance to capture the separate parts of an author's name.

4.1 Dublin Core in XML Syntax

It should be noted that the syntax used within **zetoc** and replicated in this paper is XML, but not RDF. Thus the examples in this paper should not be regarded as an exemplar for encoding Dublin Core in XML. Dublin Core element refinements are encoded as attributes of the simple element name, rather than using the more verbose 'dot' notation which would imply multiple similar elements in the XML Document Type Definition. For example, the issued date is encoded as:

```
<dc:date refine="issued">2001</dc:date>
```

rather than:

```
<dc:date.Issued>2001</dc:date.Issued>
```

Using a more recent recommendation for encoding Dublin Core in XML, which is still under development, this example would become:

```
<dc:date><dc:issued>2001</dc:issued></dc:date>
```

It should also be noted that some of the element refinements and schemes shown in the examples are not official Dublin Core qualifiers, but are taken from an application-specific **zetoc** namespace (defined at

<http://zetoc.mimas.ac.uk/zetocxx/zetocProfile>). This namespace would be specified at the head of a complete **zetoc** Dublin Core in XML record.

4.2 Bibliographic Citation Information

It is not immediately apparent how to capture the bibliographic citation elements of a **zetoc** record. These are the items indicating the article's position within a containing journal issue or conference proceedings. This bibliographic information identifies an article for citation and location purposes.

For the journal articles it was decided to follow the recommendation from the Dublin Core Citation Working Group [18] made following the 8th Dublin Core Workshop in October 2000 [19], encoded in XML. This is also the method used in other electronic journal applications developed by MIMAS [20,21]. Thus sub-elements of *dc:identifier*, qualified as a *citation*, are used within a *dccite* namespace to capture the journal title, the journal volume number, the issue or part number and the start page of the article within the journal issue. A further sub-element, within a **zetoc** namespace, is used to capture any special journal issue title, an item not considered by the DC-Citation Working Group. It should be noted that the *dccite* namespace is specific to this application and has not been ratified by Dublin Core Metadata Initiative. The ISSN of the journal is captured by *dc:relation* with an *isPartOf* refinement.

The encoding of bibliographic citation information for conference papers in Dublin Core has not yet been considered in detail by a Dublin Core working group, so a **zetoc**-specific encoding is used. Similar to the encoding for journal article citations, the information is captured by sub-elements of *dc:identifier* qualified as a *citation*, but within a **zetoc** namespace. These sub-elements can record the conference title, type and venue, and the name of the conference proceedings. The conference date is captured as a *dc:date*, qualified as *conference*, in a simple text reflecting how the data is supplied. Any additional information about the conference is captured in *dc:description*. The ISBN of the conference proceedings is included in a *dc:relation* element with an *isPartOf* qualifier.

An example **zetoc** record for a journal article is:

```
<zetocrec>
<dc:title>Dublin Core Metadata for Electronic
Journals</dc:title>
<dc:creator scheme="zetoc">
  <zetoc:snm>Apps</zetoc:snm>
  <zetoc:inits>A</zetoc:inits>
</dc:creator>
<dc:creator scheme="zetoc">
  <zetoc:snm>MacIntyre</zetoc:snm>
  <zetoc:inits>R</zetoc:inits>
</dc:creator>
<!--Library of Congress-->
```

```
<dc:subject scheme="LCC">
  TP372.5</dc:subject>
<!--Dewey -->
<dc:subject scheme="DDC">
  004</dc:subject>
<dc:publisher>
  <zetoc:pnm>Springer-Verlag</zetoc:pnm>
  <zetoc:country>Germany</zetoc:country>
</dc:publisher>
<dc:contributor scheme="zetoc" role="editor">
  <zetoc:snm>Borbinha</zetoc:snm>
  <zetoc:inits>J</zetoc:inits>
</dc:contributor>
<dc:contributor scheme="zetoc" role="editor">
  <zetoc:snm>Baker</zetoc:snm>
  <zetoc:inits>T</zetoc:inits>
</dc:contributor>
<dc:date refine="issued" scheme="W3CDTF">
  2000</dc:date>
<!--zetoc unique identifier-->
<dc:identifier refine="zetoc">
  RN085008791</dc:identifier>
<dc:identifier refine="shelfMark">
  5180.185000</dc:identifier>
<dc:identifier refine="citation">
  <dccite:journalTitleFull>Lecture Notes in
Computer Science</dccite:journalTitleFull>
  <dccite:journalVolume>1923
  </dccite:journalVolume>
  <dccite:journalPages>
    <zetoc:ppf>93</zetoc:ppf>
    <zetoc:ppl>102</zetoc:ppl>
  </dccite:journalPages>
  <zetoc:journalIssueTitle>Research and
Advanced Technology for Digital Libraries
  </zetoc:journalIssueTitle>
</dc:identifier>
<dc:language scheme="RFC1766">
  en</dc:language>
<dc:relation refine="isPartOf" scheme="ISSN">
  0302-9743</dc:relation>
</zetocrec>
```

The full **zetoc** record includes some extra fields for internal use, such as the **zetoc** record creation date, which are omitted from this example.

This particular journal article citation does not include an issue or part number, but for citations where this is necessary it would be included as a *dccite:journalIssueNumber* sub-element.

Because this article is also a conference paper, it is recorded again in **zetoc** as a conference paper. For this case, the **zetoc** record will have two instances of *dc:identifier*, qualified as *citation*, taking advantage of the fact that all Dublin Core elements are repeatable. The additional conference fields are:

```

<dc:subject scheme="keyword">
  digital libraries</dc:subject>
<dc:date refine="conference">
  2000; Sep</dc:date>
<dc:identifier refine="citation">
  <zetoc:confTitle>
    ECDL 2000</zetoc:confTitle>
  <zetoc:confType>European conference;
  4th</zetoc:confType>
  <zetoc:confVenue>Lisbon</zetoc:confVenue>
  <zetoc:proceedings>Digital Libraries
  </zetoc:proceedings>
</dc:identifier>
<dc:relation refine="isPartOf" scheme="ISBN">
  3540410236</dc:relation>

```

5 Mapping to Z39.50

In addition to a Web search interface, *zetoc* has a Z39.50 interface. It allows for searching via the Z39.50 Bib-1 Attribute Set [22], and will return information as SUTRS (both brief and full records), GRS-1 (Generic Record Syntax) and a simple tagged reference format [23]. In order to be Bath Profile compliant, *zetoc* also has the option to return Dublin Core within XML records. 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, eg. 'TI:' precedes a title, again in plain text. This format may be used for importing citations into a personal bibliographic database, and will be extended in the future to include several standard reference formats. Within the *zetoc* enhancement prototype the Z39.50 interface is provided by the enabling Cheshire software. The SUTRS and simple tagged reference formats are returned to the requesting Z39.50 client via a bespoke filter program which transforms the raw XML *zetoc* records.

5.1 Mapping to Bib-1

The indexes generated by the Cheshire software from the XML data files are mapped to Z39.50 Bib-1 attributes within the configuration file for the database. This allows a Z39.50 client to request a search on specific fields of a *zetoc* record. Some of the significant detailed mappings are shown in Table 1. This mapping is project-specific and is not presented as an official mapping from Dublin Core to Bib-1. Mapping is also included within the configuration file to the Bib-1 Dublin Core attributes which are not shown in this table. It should be noted that Bib-1 does not provide attributes for capturing article-level information such as journal volume and issue number and page numbers, and locally defined attribute values were required for these.

Table 1. *zetoc* Bib-1 to Dublin Core Mapping

Name	Code	<i>zetoc</i> DC
Conference name	3	zetoc conference sub-elements
Title	4	dc:title
Title series	5	dccite: journalTitleFull
ISBN	7	dc:relation/ISBN
ISSN	8	dc:relation/ISSN
Library of Congress classification	9	dc:subject/LCC
Dewey classification	13	dc:subject/DDC
Subject heading	21	dc:subject/Keyword
Date of Publication	31	dc:date/Issued
Code-Language	54	dc:language
Author	1003	dc:creator
Editor	1020	dc:contributor/ Editor
Country-publication	1053	zetoc:country
Date of conference	1054	dc:date/conference
Place of conference	1067	zetoc:confVenue

5.2 Mapping to GRS-1

GRS-1 (Generic Record Syntax) is a defined record retrieval syntax within the Z39.50 protocol. Mappings from the *zetoc* Dublin Core elements to GRS-1 Tagset-G elements [24] are defined in the Cheshire configuration file for the *zetoc* database and are shown in the Table 2. Cheshire uses this configuration information to return GRS-1 to a requesting Z39.50 client.

Table 2. *zetoc* GRS-1 to Dublin Core Mapping

Name	Tagset-G	<i>zetoc</i> DC
title	1	dc:title
author	2	dc:creator
publicationPlace	3	zetoc:country
dateTime	8	dc:date/issued and conference
language	20	dc:language
subject	21	dc:subject
identifier	28	dc:identifier/ ZETOC and shelfMark
publisher	31	zetoc:pnm
contributor	32	dc:contributor
source	33	citation information

5.3 Bath Profile Dublin Core

The Bath Profile of Z39.50 requires results returned as Dublin Core when a search request specifies XML. The returned XML must conform to the prescribed CIMI Document Type Definition [12] for simple Dublin Core. Within *zetoc* this XML is provided by returning search results to the requesting Z39.50

client via a bespoke filter program which translates *zetoc* XML to CIMI XML. Because *zetoc* records are held as Dublin Core, the transformation is very simple in most cases. But the problem again recurs of how to return the bibliographic citation information. Qualified Dublin Core may not be used if Bath Profile compliancy is to be retained, because the Bath Profile currently specifically prescribes the use of simple Dublin Core. This has been resolved for *zetoc* by employing a Dublin Core Structured Value (DCSV) [25] within an instance of an Identifier element to return the citation information contained within *zetoc* records. Although DCSV uses a defined syntax, making it machine parsable, it is sufficiently ‘uncryptic’ to be human readable. Another option would have been to construct a SICI [26] for the article to encode the citation information, but a SICI cannot record the journal title.

The returned Z39.50 XML search result display for the previous example record would be:

```
<dc-record >
<title>Dublin Core Metadata for Electronic
Journals</title>
<creator>Apps, A</creator>
<creator>MacIntyre, R</creator>
<subject>TP372.5</subject>
<subject>004</subject>
<publisher>
Germany : Springer-Verlag</publisher>
<contributor>Borbinha, J</contributor>
<contributor>Baker, T</contributor>
<date>2000</date>
<identifier>0302-9743</identifier><!--ISSN-->
<identifier>RN085008791</identifier>
<identifier>5180.185000</identifier>
<identifier>JournalTitleFull=Lecture Notes in
Computer Science [Research and Advanced
Technology for Digital Libraries];
Chronology=2000; JournalVolume=1923;
JournalPages=93-102
</identifier>
<language>en</language>
</dc-record>
```

Again, if a journal part number were included it would be held as a *JournalIssueNumber* within the citation identifier DCSV. The additional information which would be included for a conference paper record is:

```
<subject>digital libraries</subject>
<identifier>3540410236</identifier> <!--ISBN-->
<identifier>ConfTitle= European conference, 4th
Digital Libraries ECDL 2000; ConfVenue=Lisbon;
ConfDate=2000, Sep
</identifier>
```

It may be noted that much of the richness of the information in the *zetoc* qualified Dublin Core records has been lost.

6 Future *zetoc* Interfaces

It is expected that other standard interfaces to *zetoc* will be developed in the *zetoc* enhancement prototype. In particular, *zetoc* will provide OpenURL [27] enabled links as a step towards providing access to the full text of discovered articles for end-users, and maybe Digital Object Identifiers for the same purpose. OpenURL is an emerging standard currently undergoing NISO discussions.

6.1 Mapping to OpenURL

Within the *zetoc* enhancement prototype, a link to Articles Direct [28] at the British Library Document Supply Centre has been implemented as a ‘proof of concept’, which would enable end-user ordering of discovered articles. The link from the full search results page to this facility is enabled using the ‘Object-Metadata-Zone’ of the OpenURL protocol. The mapping from the *zetoc* data to the OpenURL fields is shown in Table 3. A more general crosswalk between Dublin Core and OpenURL is given in [29]. The link to Articles Direct is for journal articles only but the OpenURL protocol also includes a ‘conference proceeding’ genre.

Table 3. *zetoc* OpenURL Mapping

Description	OpenURL	<i>zetoc</i> DC
Record type	genre	article
Journal title	title	dccite: journalTitleFull
Article title	atitle	dc:title
First author family name	aulast	zetoc:snm for first creator
First author initials	auinit	zetoc:inits for first creator
Publication Year	date	dc:date/issued
Journal volume	vol	dccite: journalVolume
Journal issue/part number	part	dccite:journal IssueNumber
Pagination	pages	zetoc:ppf- zetoc:ppl
ISSN	issn	dc:relation / isPartOf / ISSN

An example of the metadata description part of an OpenURL is as follows. For a complete OpenURL this would follow the Base URL of a resolution service, separated from it by a ‘?’. Note that spaces have been escape-encoded to ‘%20’ for HTTP transmission. Line breaks in the example are for clarity only.

genre=article
 &title=Lecture%20Notes%20in
 %20Computer%20Science
 &title=Dublin%20Core%20Metadata%20for
 %20Electronic%20Journals
 &aulast=Apps&auinit=A
 &date=2000&vol=1923
 &pages=93-102&issn=0302-9743

The fields of the *zetoc* records shown in Table 4 are not currently included in the OpenURL specification so they are included in the article ordering URL link using the 'Local-Identifier-Zone' part of an OpenURL. The last of these, the British Library shelf-mark, is really a local identifier.

Table 4. *zetoc* local OpenURL fields

Description	<i>zetoc</i> DC
Publisher	zetoc:pnm
Country of publication	zetoc:country
Shelfmark	dc:identifier / shelfMark

7 *zetoc* Alerts

The popular *zetoc* Alert service currently sends out email tables of contents of new journal issues to requesting users. In the initial *zetoc* implementation the data feed for the alert service was the BRS-format *zetoc* update file. This data feed has now been changed to an XML file with Dublin Core *zetoc* records, as for the *zetoc* enhancement prototype database. The alert email messages are currently in plain text. Changing the alert data feed into an open standard format opens up possibilities of offering *zetoc* alerts in several standard formats such as XML, Dublin Core, and RDF Site Summary (RSS) [30], as well as tagged bibliographic formats. Providing alerts in RSS would enable their use for news feeds, whereas tagged bibliographic formats may be imported directly into personal bibliographic databases. It is also planned to provide subject-based alerts, implemented via a simple keyword search on the *zetoc* update data, which again could be offered in several standard formats.

8 Interoperability Issues

8.1 Standard Interface Formats

It is not strictly necessary for the data within *zetoc* to use Dublin Core or to adhere to an open standard, though it seems good practice to encode data in a standard way. For the *zetoc* Web display, the internal *zetoc* data records are converted to HTML, so the format of the base data is irrelevant. The same is true for some of the Z39.50 formats provided, where the internal *zetoc* data is filtered before being returned to the requesting Z39.50 client. Thus the internal *zetoc* data encoding could have used element names

different from Dublin Core, with appropriate transformation used for data delivery in most cases. But it was decided that, as well as reinforcing good practice, holding the data in Dublin Core would simplify these and any future data transformations. However, standard data formats are required for interoperability where open standard interfaces are used. This is the case for the Z39.50 GRS-1 and XML interfaces.

8.2 Dublin Core for Resource Discovery

There is a significant body of opinion within the Dublin Core community that Dublin Core should be used primarily for simple resource discovery [31], thus making the definition of compound element values undesirable. It is certainly true that the take up of Dublin Core has been aided by its simplicity of concept for everyone, rather than being just for subject specialists. When used for resource discovery, whether by general searching over the World Wide Web, or using more specific resource discovery services such as *zetoc*, it would seem essential that a human-readable record be returned. *zetoc* provides textual brief and full search results through its Web interface and the SUTRS Z39.50 format. But this does not mean that all fields of the record which do not fit obviously into a Dublin Core element should be included within an unstructured Description element, because there may also be a requirement, in the future if not now, for further machine processing of the returned record.

8.3 Dublin Core for Resource Description

Although Dublin Core was originally conceived for resource discovery it is increasingly being used for resource description. It is necessary to balance the desirability of maintaining the simplicity of Dublin Core against the wish for more complexity to capture information about real systems. Dublin Core should remain a 'core' set of metadata elements, with domain-specific metadata recorded according to more complex standards, whether extensions to Dublin Core or separate standards. For instance, replicating a full library catalogue within simple Dublin Core elements would not necessarily be an acceptable use for Dublin Core. However, the bibliographic citation of a journal article seems to be fairly fundamental information, required within many subject areas, at least for academia and researchers. How to capture such citation information is a problem which many people have already encountered when trying to use Dublin Core for resource description. Thus it would seem sensible to have a recommended best practice method for capturing this information in Dublin Core.

8.4 Metadata for Citations

There is currently no mechanism formally recommended by the Dublin Core Metadata Initiative for encoding bibliographic citation within the Dublin Core Element Set. The Dublin Core Citation Working Group has discussed the capture of bibliographic citations for journal articles [29]. This is still an open issue, though they are likely to suggest a recommendation similar to that used within *zetoc* for journal articles, encoding sub-elements in some way within *dc:identifier*. Using the Identifier element recognises the fact that the set of citation information effectively *identifies* the article, and could be used for discovery of the indicated full article.

The Dublin Core community has not yet investigated encoding bibliographic citations for other genre. Possibly recommendations for conference papers, book chapters and other scholarly literature will become work items for a future working group. Some other metadata initiatives have made recommendations in this area. OpenURL includes metadata for conference papers. A standard specification for conference title pages has recently been published by NISO [32] which should aid the standardisation of metadata in this area.

8.5 Citations within Simple Dublin Core

As indicated above the requirement for Z39.50 Bath Profile interoperability raises the question of how to provide the bibliographic citation information within *zetoc* in an interoperable way through the prescribed simple Dublin Core XML format. *zetoc* has chosen to provide this information via the Z39.50 interface using a Dublin Core Structured Value (DCSV) within an instance of the Identifier element. DCSV is a ratified syntax, but there are as yet no recommendations for the labels for the citation 'sub-elements' within this DCSV, making the interoperability of this approach questionable.

8.6 Hierarchical Metadata

Most of the information required for a journal article citation, as opposed to the information about the article itself such as its title and authors, is information about the containing journal volume and issue. The exception to this is the pagination information which records the location of the article within the printed version of the journal issue and thus is pertinent to the particular article. In future, and some current, electronic journal publishing models, this pagination information will become irrelevant, though would by necessity be replaced by some other numbering. But, at the present time, recording the position of an article within a printed journal is the generally used model and a requirement for reference linking.

Some may argue that information about the journal issue should be pointed to from the article's metadata, for instance using a *dc:relation* element with an *isPartOf* refinement, and likewise metadata for the journal issue or volume should point to metadata about the journal itself. This mechanism is in fact used in *zetoc* to record the ISSN of a journal. Theoretically this approach is correct, but it is probably not viable in all practical environments. Within a current awareness application like *zetoc*, all the information about the article including its citation, which records its whereabouts within a journal issue, is held in one record, with little knowledge of, or ability to access, information about the journal. The end-user will expect to see all the information about a discovered article within one search result.

8.7 Application Specific Schema

It has been suggested that the bibliographic citation for an article is application specific information, and so should be captured within application specific elements and qualifiers according to an application profile [33]. To some extent this approach has been explored within *zetoc*. Information which is specific to *zetoc*, such as identifiers and location codes, is recorded according to schemes within a *zetoc* namespace. However, capturing bibliographic citation information seems to be a more generic, cross-domain problem. It is information which is becoming increasingly significant with the implementation of linking technologies [34] and the requirement to be able to locate appropriate copies of articles for end-users [14]. Possibly the mechanism for recording this information within a *dc:identifier* element should become part of a 'citation profile' but it appears to be a general enough requirement for it to become Dublin Core best practice. Whether a citation is a sufficiently generally used mechanism to merit a new element within a 'citation' namespace, or whether there should be a more general, hierarchical 'container' element within Dublin Core are open questions.

9 Conclusion

The *zetoc* current awareness service has provided a platform, with a significant amount of data, to investigate mechanisms for capturing journal article and conference paper records using the Dublin Core metadata element set, and displaying such records as discovered search results. Although the use of Dublin Core for encoding the internal *zetoc* data was not strictly necessary it has highlighted areas where Dublin Core mechanisms and best practice recommendations would assist resource description and hence resource discovery, location and acquisition. In particular, recommendations are seen

to be lacking in the area of metadata for the bibliographic citation of journal articles and conference papers.

zetoc has also provided a case study to explore interoperability between several open standard formats, in particular between Dublin Core and some of the Z39.50 attribute and syntax codes, within a service environment.

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) [35] for UK higher and further education, as part of the Join-Up programme [36] within the Distributed National Electronic Resource (DNER) development programme [37].

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] Z39.50, the North American National Information Standards Organisation (NISO) standard for information retrieval. <http://www.niso.org/z3950.html>
- [4] P. Miller. (1999) Z39.50 for All. *Ariadne* 21, September 1999. <http://www.ariadne.ac.uk/issue21/z3950>
- [5] The Z39.50 Bath Profile. <http://www.nlc-bnc.ca/bath/bp-current.htm>
- [6] The Dublin Core Metadata Initiative (DCMI). <http://www.dublincore.org>
- [7] The Dublin Core Metadata Element Set. <http://www.dublincore.org/documents/dces>
- [8] XML (eXtensible Markup Language). <http://www.w3.org/XML>
- [9] Electronic Publishing at MIMAS, a UK Higher and Further Education data centre. <http://epub.mimas.ac.uk>
- [10] The COPAC research library online catalogue service. <http://copac.ac.uk>
- [11] BRS/Search. <http://www.opentext.com/dataware/>
- [12] 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>
- [13] The Cheshire II Information Retrieval System. <http://cheshire.lib.berkeley.edu>
- [14] H. Van de Sompel and O. Beit-Arie. Open Linking in the Scholarly Information Environment Using the OpenURL Framework. *D-Lib Magazine* 7(3), March 2001. doi://10.1045/march-vandesompel
- [15] 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
- [16] CrossRef. <http://www.crossref.org>
- [17] Dublin Core Qualifiers. <http://www.dublincore.org/documents/dcmes-qualifiers/>
- [18] The DCMI Bibliographic Citations Working Group. <http://www.dublincore.org/groups/citation/>
- [19] S. L. Weibel and T. Koch. The Dublin Core Metadata Initiative Mission, Current Activities, and Future Directions. *D-Lib Magazine* 6(12), December 2000. doi://10.1045/december2000-weibel
- [20] A. Apps and R. MacIntyre. Dublin Core Metadata for Electronic Journals. *Proceedings of the 4th European Conference on Research and Advanced Technology for Digital Libraries (ECDL2000)*, Lisbon, Portugal, 18-20 September 2000. *Lecture Notes in Computer Science* (Springer-Verlag) 1923:93-102, 2000.
- [21] A. Apps and R. MacIntyre. CABRef: Cross-Referencing into an Abstracts Database. *Proceedings of the ICC/IFIP 5th Conference on Electronic Publishing (ELPUB01): "2001 in the Digital Publishing Odyssey"*, Canterbury, UK, 5-7 July 2001. (IOS Press), pp 256-265, 2001. <http://epub.mimas.ac.uk/papers/appsmacep2001.html>
- [22] The Z39.50 Bib-1 Attribute Set. <http://lcweb.loc.gov/z3950/agency/defns/bib1.html>
- [23] Accessing **zetoc** with a Z39.50 client. <http://zetoc.mimas.ac.uk/z3950.html>
- [24] The Z39.50 Generic Record Syntax (GRS-1) Tagsets. <http://lcweb.loc.gov/z3950/agency/defns/tag-gm.html>
- [25] S. Cox and R. Ianella. DCMI DCSV: A syntax for writing a list of labelled values in a text string. 2000. <http://www.dublincore.org/documents/dcmi-dcsv/>
- [26] SICI, Serial Item and Contribution Identifier. <http://sunsite.berkeley.edu/SICI>
- [27] OpenURL Syntax Description. <http://www.sfxit.com/OpenURL/openurl.html>
- [28] Articles Direct, the British Library Document Supply Centre. <http://www.bl.uk/services/bsds/dsc/artdir01.html>
- [29] A. Powell and A. Apps. Encoding OpenURLs in Dublin Core metadata. *Ariadne* 27, March 2001. <http://ariadne.ac.uk/issue27/metadata>
- [30] RDF Site Summary (RSS). <http://purl.org/rss/1.0/spec>
- [31] C. Lagoze. Keeping Dublin Core Simple. *D-Lib Magazine* 7(1), January 2001. doi://10.1045/january2001-lagoze
- [32] NISO Z39.82-2001. Title Pages for Conference Publications. <http://www.niso.org/commitao.html>
- [33] R. Heery, and M. Patel. Application profiles: mixing and matching metadata schemas. *Ariadne* 25, September 2000. <http://www.ariadne.ac.uk/issue25/app-profiles>
- [34] P. Caplan and W. Y. Arms. Reference Linking for Journal Articles. *D-Lib Magazine* 5(7/8), July/August 1999. doi://10.1045/july99-caplan
- [35] The Joint Information Systems Committee (JISC). <http://www.jisc.ac.uk>
- [36] The Join-Up programme. <http://edina.ed.ac.uk/projects/joinup/>
- [37] The UK Distributed National Electronic Resource (DNER). <http://www.jisc.ac.uk/dner/>