

Cataloguing the World Wide Web: CORC at Edinburgh University

Background

Having successfully operated the WorldCatⁱ co-operative cataloguing model for many years OCLCⁱⁱ decided to explore using the same concept to ease the burden and duplication of effort of individual libraries organising, describing and presenting web resources for their users. However, rather than encourage libraries to add web resources to WorldCat OCLC decided to design a system which would incorporate the benefits of co-operative cataloguing with an automatic metadata harvesting or extraction tool and the flexibility of being able to map between multiple metadata formats. The Co-operative Online Resource Catalogue (CORC)ⁱⁱⁱ was thus conceived as a co-operative system which would also help to automate the record creation process and allow libraries to utilise different metadata schemes to suit their requirements.

OCLC sent out a call for participants to test out the system in late 1998 and released the first version of CORC early in 1999. Records were initially loaded into the CORC database from OCLC's InterCat^{iv} and NetFirst^v databases. Participants were asked to commit 0.5 FTE to use the system to search for, edit and add new resources. They were also encouraged to provide active feedback which was used to develop the system with regular releases over the period of the project. OCLC also hosted a number of participants' meetings to further garner ideas about and problems encountered with the system. During this period CORC was made available free of charge to all participating institutions.

How CORC works

CORC provides a 'button' link which can be dragged from the system and dropped directly into the user's personal toolbar. The user simply clicks on this button when viewing a site of interest on the web browser and CORC automatically starts the record creation process by harvesting metadata from the HTML metatags. The user is taken directly into CORC and presented with a 'raw' metadata record for the chosen web site. Alternatively, it is possible to enter CORC through its homepage, choose the record creation area and paste in the URL of the site. Multiple records can be created by this method; the user specifies the URL of a site and also the number of links on that site that they would like to create records for. CORC then creates basic records for these sites and informs the user when they are available for editing.

Once inside the record creation process the editor has access to the record and the web site it describes, which is displayed on a lower frame of the page. This permits the editor to refer to the web site and edit the record accordingly. If preferred the whole screen can be dedicated to the record and the Web site viewed on a separate browser window. The automatically generated fields can be edited using the normal cut-and-paste functions and other fields are deleted and added as required with the click of a button. In response to concern from cataloguers CORC added alternative editing modes in addition to their original template options which require browser interaction every time a field is added or deleted. Cataloguers felt this made the editing process unacceptably slow so a text area was developed which allowed users to edit records with fewer browser interactions, so speeding up the process.

Sets of constant data can be created, stored and used when editing a group of records displaying significant similarities. Constant data can be automatically added at the initial record creation stage by CORC or at any later stage by the cataloguer from a list of actions on a drop down menu.

CORC also trialed the WebDewey^{vi} product with the project phase of CORC. This automatically creates Dewey classification numbers for web sites some of which are then linked to appropriate Library of

Congress Subject Headings. Unfortunately, experience at Edinburgh suggested that many of the class numbers generated were very wide of the mark and as the Library had recently moved over to the Library of Congress classification scheme it was decided not to proceed with this tool when it became a separate chargeable product.

Authority Control

CORC offers cataloguers the option to control all the headings in a record in one step or on an individual basis. Headings are checked with an authority control button which either accepts the heading as valid or opens up an authority box suggesting alternatives and allowing the cataloguer to search the OCLC authority files. Valid headings become highlighted links which lead to the authority records. It is also possible to search and browse the authority indexes for suitable headings.

Saving and Exporting Records

Cataloguers can either save records into the main Resource Catalogue where they will be accessible to all CORC users or they can choose to keep them private by using the Savefile area. CORC provides each institution with a Savefile area which can be used in several ways. It can store new records which are still works in progress. Once new records are completed they can be moved into the main database making them accessible to all participants and indeed users must move new records into the public domain before they can be exported. The Savefile can also be used to store records which have been copied or cloned from the main database for the purpose of local editing before records are exported. CORC also allows some restricted editing of 'master' records on the main database so that inaccuracies, such as invalid URLs, can be corrected.

CORC automatically validates records and warns cataloguers of any MARC formatting errors before they are submitted to the main database or exported to another system.

Records can be exported singly or in a batch from CORC in either MARC or Dublin Core HTML or RDF formats.

CORC has also introduced a link checking service. Users can check periodically on CORC for a list of invalid or redirected URLs the system has discovered. Users are only informed about URLs on records they have either created or have requested to be informed about – such as records they have exported. It is then up to the user to correct the URL on CORC and their local system as necessary.

Metadata Formats

CORC supports both MARC21 and Dublin Core formats. Work at Edinburgh University has concentrated mainly on using MARC so that records for web resources can be imported into the Library's Endeavor Voyager catalogue. CORC provides context sensitive help and support with MARC editing, with links from each MARC tag and links via the home page to other information sources such as Nancy Olson's Guide to Cataloguing Internet Resources^{vii}.

Dublin Core users have the option of following 'simple' Dublin Core or using qualifiers as developed by the DCMI^{viii}. CORC has also developed additional qualifiers which reflect library users' experience with MARC and the desire to be able to map easily between the two formats. Examples of CORC qualifiers include the addition of *Personal*, *Corporate* and *Conference* with the *Contributor* element and the qualifier *is Part of Series* with the *Relation* element. CORC also provides online help with editing in Dublin Core.

It is possible to move between the two formats at any time when viewing or editing a record. Not all information will be visible in both formats as some fields in one format do not have an equivalent in the other, however no information should be lost in the mapping and all will reappear if the record is subsequently viewed in the original format. Mapping from MARC to Dublin Core tends to lead to a more satisfactory result than the reverse action due to the strict rules and guidelines associated with MARC.

Pathfinders

In addition to the record creation side of the system, CORC also has another related strand in the form of subject bibliographies or 'pathfinders'. CORC aims to help libraries by automating the process of creating subject gateways to both digital and physical resources and providing access to its database of web resources. Benefits also include the link maintenance mentioned earlier and the dynamic search feature which allows pathfinders to be automatically updated as new resources are added to the main catalogue. As with resource records libraries can also benefit from co-operative effort as it is possible to copy, edit and export pathfinders so saving on the time and effort of creating resource pages from scratch. Pathfinders are a valuable function for libraries wishing to set up pages of resource links within specific subject domains quickly. In practice however, particularly in academic libraries, it is likely that subject librarians would find the value of them to lie mainly as an alerting tool to new web resources in a domain, which can then prompt further editing of already existing pages produced locally.

Pathfinders have yet to be used at Edinburgh, but they are currently being evaluated as a potential tool for updating and enhancing the Library's current subject web pages.

Future of CORC

CORC was launched as a fully chargeable product in July 2000, following a similar usage and pricing structure as WorldCat. Both services have been integrated, with all records submitted to the main catalogue in CORC simultaneously added to WorldCat and any records containing 856 links saved to WorldCat uploaded to CORC on a daily basis. WorldCat can be searched from CORC, but it is not possible to search for 'web only' resources from WorldCat.

CORC is not a static product and OCLC is still looking at ways of improving the service with enhancements added to each new release. Areas of potential future development include the incorporation of other metadata formats such as the IMS^{ix} metadata standard. It is also hoped that CORC's functionality might be extended to aid with the problems associated with digital preservation - in particular the issue of archiving web sites.**CORC at Edinburgh University**

Through the Science and Engineering Library and Learning Information Centre (SELLIC) project, Edinburgh University Library joined the project in October 1999 after the appointment of a Metadata Editor. Transatlantic training was arranged by way of an audio conference and the Metadata Editor introduced to the system. The initial training was followed up with a short period of time using the CORC practice system before the Metadata Editor started to add resources to the main database.

CORC has been used primarily to create MARC records for electronic resources including departmental web sites, electronic abstracting and indexing databases and for web resources recommended by academics on course web sites. The Library is also currently considering changing its current electronic journal policy of simply adding 856 links to records for print journals to one of creating separate records for electronic journals. CORC would be useful here as a repository of records which could be easily edited and imported into the Library's catalogue.

The Library is also looking at ways of encouraging academic staff to recommend web sites for the

catalogue in much the same way as they recommend print resources. The University's Web Editor is developing a simple web form which will allow academics to simply paste in a URL which will then be sent to a database of URLs so recreating for the world of the web the familiar cataloguing backlog for the Metadata Editor to work from.

Finally the Library would like to encourage the incorporation of good quality metadata in the University's own web pages, enriching them and making them potentially more useful to the science and engineering communities worldwide. CORC is one potential tool for creating Dublin Core HTML metatags which could be easily added into the HTML source for these pages.

Conclusion

The University of Edinburgh's experience with the metadata side of CORC has been largely positive. Access to an expanding database of good quality records for web resources saves the Library time and effort in creating records from scratch. The system saves on some original cataloguing with the automatic creation of a basic record, and allows the cataloguer conveniently to view the web site and record all on one screen. It provides good online help for editing in both MARC and Dublin Core, automatically maps between the two formats, links to OCLC authority files and provides URL checking. In this respect it is a useful tool which allows the Library to describe, present and maintain selected web resources for its users. However, on the downside, the raw metadata records are very basic and require a considerable amount of editing to bring up to an acceptable standard, the editing process tends to be somewhat slower than with a traditional library system cataloguing module due to the reliance on the web and the variable speed of browser interactions, and finally there is still a considerable degree of US bias with regard to the resources available on the system. We would wish to see this addressed through much greater adoption of the service by European academic libraries.

Zena Mulligan
SELLIC Metadata Editor
Edinburgh University Library

John MacColl
SELLIC Director & Sub-Librarian, Online Services
Edinburgh University Library

ⁱ OCLC *WorldCat*: <http://www.oclc.org/oclc/menu/colpro.htm>

ⁱⁱ OCLC: <http://www.oclc.org/home/>

ⁱⁱⁱ OCLC *Co-operative online resource catalogue*: <http://www.oclc.org/oclc/corc/> last accessed 04.04.01

^{iv} OCLC *InterCAT project*: <http://www.oclc.org/oclc/research/projects/intercat.htm>

^v OCLC *NetFirst*: <http://www.oclc.org/oclc/netfirst/>

^{vi} OCLC *Dewey Decimal Classification: WebDewey in CORC*:

<http://www.oclc.org/oclc/fp/products/webdeweyincorc/webdeweyincorc.htm>

^{vii} Olson, Nancy ed. *Cataloguing internet resources: a manual and practical guide*, 2nd ed. OCLC, 1997:

<http://www.oclc.org/oclc/man/9256cat/toc.htm>

^{viii} Dublin Core Metadata Initiative: <http://dublincore.org/>

^{ix} IMS Global Learning Consortium, Inc.: <http://www.imsproject.org/>