

From INTERCAT to CORC:
Cooperation in the selection and cataloguing of Web resources

by

Stuart Hunt
OCLC Europe, the Middle East & Africa

The purpose of this paper is to highlight the achievements of two OCLC projects. One of these, a research project, INTERCAT, contributed significantly to the enhancement of already existing OCLC systems and services by its outcomes. The other, the Cooperative Online Resource Catalog, or CORC, led to the establishment of what was an entirely new OCLC service.

Through the discussion of these two projects I shall draw out what, I believe, are the lessons learned for the co-operative model of the bibliographic utility and the individual library. Particular attention will be paid to the concepts of co-operation inherent in each of these projects. I shall aim to highlight the implications for both the bibliographic control and the collection development practices of libraries with regard to internet resources. Firstly, by way of scene-setting, I will begin with some thoughts on co-operation.

Co-operation is the central tenet of the bibliographic utility. The model of shared co-operative cataloguing has been, and continues to be, widely adopted throughout the library community. Changes in the economic circumstances for libraries, combined with the benefits of co-operation are themes that never lose favour for libraries. The rise and strength of the consortium is our current embodiment of this. As metadata workers of one sort or another we recognise that the concept of co-operation grounds much of our work. What is interoperability but another means of expressing this? What, today, is different for us is that which defines our term 'co-operation'. Whereas, until quite recently libraries saw co-operation in terms of relationships with other libraries, now our co-operation is much extended. To this extent we can say that much has been achieved thus far, that our libraries already work together across national and cultural boundaries and that this brings benefit to all.

In our globally networked environment we recognise that co-operation extends well beyond our libraries into other sectors and domains. Much discussion has taken, and continues to take, place at the cross-sectoral level, leading to new and challenging models of co-operation. We sit on committees, working groups, and attend conferences and meetings where it is a given that we are no longer exclusively library workers. The metadata schemas that we work and inter-work with are a graphic illustration of the new co-operation that the web has engendered in us.

The path that has led us to this place is one of many steps and each domain may be able to point to key historical events that truly brought a paradigm shift to our thinking and, more importantly, our practice. For OCLC, and its member libraries, the first real steps began in the early 1990's.

INTERCAT

During the 1990's OCLC's Office of Research began investigating the cataloguing of internet resources. This activity was made carried out with the assistance of two U.S. Department of Education grants. The "Internet Resources Project" (1991-1993) was central to the conceptualisation of the recording of the remote location of an internet resource. The fundamental outcome of this research was the move towards a USMARC tag for the recording of both the electronic location and access method for internet resources. This, as we all know, came with the creation of the 856 USMARC (as it then was) tag.

The 'Internet Resources Project' was followed by another project entitled "Building a Catalog of Internet Resources" (1993-1996), what is better know as the INTERCAT project. The specific focus of INTERCAT was the creation of MARC bibliographic data for internet resources. Building on the vision of the 'Internet Resources Project', INTERCAT participant institutions cooperated in the creation of a pool database of bibliographic records for internet-only resources.

Concurrent with INTERCAT was the approval by the MARBI (Machine-Readable Bibliographic Information) 1994 committee of the use of the 856 USMARC. INTERCAT was thus given a significant boost by the validation of its practice within the MARC community. Libraries were thus, for the first time, officially able to record the remote location of electronic resources within the MARC bibliographic record in a standardised manner.

The over-riding objective of the INTERCAT project was to assess the use of (what was then) the USMARC format for bibliographic data in order to provide description, location, and access information for remotely accessible electronic resources. Specifically the 856 tag for electronic location and access would be used. The project was designed with the following aims in mind:

- The creation of a database of USMARC bibliographic records for remote access electronic files through a coordinated, cooperative effort
- The provision of wide-scale access to this catalogue through WorldCat, the OCLC Online Union Catalog, OCLC's FirstSearch service, and a specially created database freely accessible to all Internet users
- To complete the link between coded location and access data in the bibliographic surrogate (via the 856 tag) and the objects themselves.ⁱ

INTERCAT had specific project measurables in terms of the creation of a shared the database of resources, and in providing a support infrastructure for cataloguers on what was then the "leading edge". Similarly the project aimed to furnish tangible results that could be seen in the following terms:

- Aid the integration, description and access of internet resources within library catalogues
- Test the usefulness of MARC coded bibliographic data in providing effective retrieval within a searchable catalogue
- Test the technical feasibility of providing direct access from surrogate to electronic resource based on coded access and location information provided in the 856 tag
- Measure the stability of location and access information over time through repeated, automated tests to retrieve objects based on coded data in the 856 tagⁱⁱ

As host institution OCLC provided and supported the database of bibliographic records providing description, location, and access information for internet resources. A significant outcome took the form of documentation to make the cataloguing of electronic resources easier. This was realised with Cataloging internet resources: a manual and practical guide edited by Nancy Olsonⁱⁱⁱ

The INTERCAT project could be evaluated in a number of ways. It is clear that INTERCAT implied the implementation of solutions that would have fundamental implications for library practice. To this extent we can see that INTERCAT was radical in terms of the challenges it posed libraries, but probably less radical than the challenge posed by the internet itself. Taking as a given the need to bring internet resources under bibliographic control, how the library achieves this is central. Library policies would, in all likelihood, be impacted in the areas of staffing and workflows (with the decisions, for example, of how this work should be done and by whom). Likewise fundamental technical infrastructure questions as to how the library could integrate resources from outside along into collections would be effected.

At the time of the INTERCAT project there were significant advantages for libraries, operating in the integrated library system environment, to using MARC. Metadata in the form of MARC bibliographic data could be subject to a system of formal control through bibliographic description and authority control. Integrated library systems are well established in the processing of large volumes of just this type of metadata, in creating indexes, and providing effective. The end result is that all an institution's collections could be managed within one central system.

Clear positive outcomes from INTERCAT could be seen in the successful creation of a significant number of bibliographic records for internet resources. These records were created by well over two hundred libraries during the project. The records found their way into different OCLC services with all the implications for resource discovery that this implies. Each of the records contained the 856 tag, and were integrated into WorldCat, thus becoming available both to all OCLC cataloguing libraries and to library end-users via the FirstSearch service. INTERCAT showed that there was significant benefit in subjecting internet resources to MARC bibliographic control and in bringing these resources into library catalogues as just one other type of material. However, we should not forget that other events outside of INTERCAT continued to move at a rapid pace.

CORC

MARC cataloguing at the time of INTERCAT seemed like if not the best, then one of the best ways to make progress in this area. We should not ignore, however, the concurrent emergence of other metadata schemas arising out of different communities. The Dublin Core Metadata Initiative and the issues it sought (and continues to seek) to address suggests an alternative approach to what is, at some level, fundamentally the same problem: Resource discovery on the web.

Because of this parallel emergence of multiple metadata schemas and/or communities beginning to express their views and needs, there is a clear shift in the approach taken by CORC from its very inception. INTERCAT was successful in its task of cooperation. That cooperation took place between libraries in what was, essentially, the same paradigm as shared cooperative cataloguing has existed since the early 1970's: Libraries pull together, work together, to ease the load and resource share in the broadest manner.

At the start of 1999 OCLC launched the Cooperative Online Resource Catalog (CORC) as a research project, again organised out of the OCLC Office of Research. Again, one of the issues was to work on the cataloguing of web resources, but this time to also provide a toolkit to libraries to ease this process. There would be less emphasis placed on the format-specific central database as a repository of MARC bibliographic data. The CORC Resource Catalog, since day one, has also supported Dublin Core metadata and users are able to select the metadata appropriate to their needs, rather than being confined to MARC only. The existence of a MARC-Dublin Core crosswalk makes it possible to convert metadata on-the-fly.

The influence of the Dublin Core community, with the possibility of the resource creator also creating the metadata for the resource, can also be detected in the CORC toolkit. This is seen in a toolkit which allows for the automatic creation of metadata. Thus the creation of metadata can be broadened, widening the focus of co-operation to the creator as well as the collector of the resource.

During the period of CORC as a research project a clear indication of its timeliness could be seen in the number of participants. Starting with an initial aim of one hundred, the research project numbered over four hundred at its close. Significant amongst these, and unlike INTERCAT, was a number of non-US institutions. Worthy of note are those participants for whom MARC21 and AACR cataloguing was not an accepted standard and for whom the application of Dublin Core metadata offered a point of access. From an OCLC perspective success was marked by the July 2000 release of CORC as an OCLC product, thus increasing the number of libraries who may wish to co-operate in the creation of metadata for web resources.

To assess the effectiveness of CORC the issue is more complex than that of INTERCAT. CORC uses web-browser technology to interface to the service. Technical development is thus, to some extent, constrained by the functionality of the browser software. This is most apparent in a non-English language context. Whilst CORC is Unicode compatible, OCLC has tied the Resource Catalog to WorldCat, thus restricting the extent of the Unicode repertoire which can be used. This will only change with OCLC's move to a new database platform, Oracle, for WorldCat. This technical constraint has implications for co-operation as users, particularly in a European context, have been frustrated. Further complications arise with differing behaviours of web browsers. The implication of this is that, whilst co-operation is increased in CORC outside of the traditional US-centric OCLC membership, it is constrained by technology.

Co-operation

I would now like to give some further consideration to the types of co-operation inherent in each of these two projects.

INTERCAT can be characterised by inter-library co-operation; all co-operation taking place amongst libraries. The central nexus for this sharing was OCLC, organising and administering from the centre. Co-operation was achieved through the sharing of bibliographic data – very much a 'bread and butter' issue for OCLC. Essentially there is little difference in the type of metadata being shared as the model of co-operation that it is founded upon is a well-established model of the bibliographic network. MARC data is created and re-used by co-operative participants, all of whom are libraries. The only qualification to this is that, with INTERCAT, for the first time libraries were sharing electronic location and access information within the metadata.

CORC continues this model of co-operation and all of the above characteristics can be found in the use of CORC both as a research project and, now, as an OCLC product. The CORC Resource Catalog is, at this level, 'business as usual' for inter-library co-operation. However, the co-operation that CORC potentially fosters does not stop at this point. Because the CORC toolkit engenders a significant degree of ease in the creation of metadata, the use of the toolkit does not need to rest in the hands of the cataloguers. Whereas INTERCAT required highly skilled cataloguers creating MARC bibliographic data, CORC does not impose the same constraints. Selection of resources is a more visible and up-front activity in CORC. The software makes it possible for the selector to select those resources which are to be brought under some sort of bibliographic control, creating a

scratch record through automated harvesting routines, which the metadata specialist can then turn into the required quality description that the institution needs.

Who is selecting the electronic resources which are to be brought under bibliographic control within the library? The answer to this is, in all probability, not the cataloguer. Selection of resources for the library collection(s) is done by the subject specialist or selector. This model is being visibly extended through virtual collection development, as libraries take account of those web-based resources they would like to point users toward.

What we are observing is that CORC is being used by both metadata and collection development staff within libraries. These groups, historically kept apart within the library at a functional level, are working together and thus co-operating. Thus CORC engenders an additional type of co-operation to inter-library co-operation, this time what we could term intra-library co-operation. Very real practical work-flow issues have been, or are being, hammered out within the individual institution in order to realise a practical intra-library co-operation. The implications of this should not be dismissed in terms of library management and organisational structures, in particular and there are advantages and disadvantages in this process. The old models of the functional organisational structure divides staff by their skills and groups tasks by their similarities. This fosters a culture of specialisation. However, as has been pointed out^{iv} this can also run counter to the aims of co-operation as team goals assert themselves over library goals, and inter-team rivalries are possibly created. Some have argued that only by integrating what have historically been the over-specialised technical services activities of the library with the end-user oriented reader-services activities can libraries effectively meet the challenges that are posed by the digital environment^v.

CORC carries the possibility of holding a mirror to the hybrid library. The wealth of electronic resources and how the library brings these under control are issues which are creating ambiguity at least, or even the dissolution of traditional boundaries within the functionally organised library. The web, and technology, will only hasten this dissolution. CORC can embody it, highlight it, reflect it.

With CORC co-operation, then, is inter-library and intra-library. But this co-operation does not stop there. Just as the hybridisation of the library has the potential to refashion relationships, so too does the breaking-down of boundaries inherent in the hybrid environment. And in the breaking-down of these boundaries we should re-introduce the emergence of the proliferation of metadata schemas and the communities they represent.

With INTERCAT we saw the library community doing one of the things it does best in the form of co-operation. But we also saw that this was inter-library co-operation. Whilst the world outside may have been moving towards multiple metadata schemas, INTERCAT was, by the constraint of its time, MARC-centric. Coming later, and fertilised by a Dublin Core influence, CORC supports both MARC21 and Dublin Core and carries the possibility of supporting additional schemes. In this way the potential of CORC to expand the notion of co-operation is realisable. Instead of inter-library we can now think inter-community co-operation. The MARC world has achieved the inter-library aspects, now new schemas and crosswalks between schemas is achieving the inter- we hear often: Interoperability.

As the boundaries of the hybrid library are less defined so, too, are the boundaries of academic research and publication – those things libraries have needed to take account of. The explosion of scholarship on the web positions both researcher and library within this complex web of relations. In such a de-centred environment the library needs to effectively take account of changing resources and new relationships. Metadata that aids and abets in resource discovery may or may not derive from the library or the library community. Co-operation is truly inter-community.

Co-operation in the creation of metadata that raises some fundamental questions. Bibliographic utilities, of which OCLC is but one, have always operated with a central tenet of trust amongst participant members. A library has trust in the metadata created by fellow participants. At the centre, quality measures are maintained by OCLC, ensuring that the integrity of the shared database is not compromised. This is well established behaviour and both automated and manual routines ensure the smooth running of quality control measures. When we move further out into the web environment these two same issues of trust and quality become more pronounced. The democratisation of information that the web embodies multiplied the number of potential sources of metadata. Trust is a problem not only with regard to the resource itself but also to the metadata associated to the resource. Validation of the source of the metadata needs to be possible, perhaps through authentication or rights management models.^{vi}

Virtual collection development

I would now like to consider the implications for what has become known as virtual collection development.

During the early to mid 1990's there was a strong fundamental scepticism about the need for the bibliographic control of electronic resources. Questions such as should a library 'collect' these resources and what does 'collection' mean in the virtual environment as well as very real concerns over resource stability and persistence.

The statistics that show the growth of web content clearly illustrate the scale of the problem. The debates and concerns of the period have now been outweighed by the realities. Libraries are pointing users to web resources. The ways in which they do this are not uniform, but do show a remarkable degree of similarity. The resources they are guiding users toward likewise show a remarkable degree of similarity.

Libraries have formulated or are formulating virtual collection development policies. The principles that govern these policies are often revolve around the following:

- To expand the reach of library collections beyond the physical
- To foster awareness of useful web-based resources
- To add value to resources retrievable by other means

Collection development of free-of-charge electronic resources may be a relatively painless procedure for the library as it (potentially) by-passes the financial issue. Where the resource is available on a for-fee basis there may be stronger reasons to have the resource under bibliographic control in order to maximise use.

A typical virtual collection development policy, aside from requiring that a resource conform with general collection development criteria, may also carry format-specific requirements.

- Currency of the resource
- Value-added benefits the resource brings to library collections
- Compatibility with existing hardware and software infrastructure
- Ease of use for the end-user
- Rights management issues

Once an electronic resource has been selected there remain format-specific issues for the library peculiar to web-based materials. Resource stability or volatility, as well as currency, and continued appropriateness may need re-visiting on a regular basis. As the dedicated gardener tends the plants, so the virtual selector needs to tend to the electronic resources.

For resources changing over time there are bibliographic control issues and these issues are akin to those experienced by the serials cataloguer. Current debates about seriality have much to contribute. Metadata may require regular updating if there is to be a continued relevance and meaningful discoverability of the resource.

The activity of virtual collection development may be a difficult one to perform. Identifying a resource suitable for 'collection' is no easy task. The volume of web documents carries disadvantages as well as advantages. The proverbial needle and haystack is a readily applicable metaphor.

Having said this, both INTERCAT and CORC do present a workable model of how the selection of resources can be made easier. The INTERCAT database and the CORC Resource Catalog contain metadata for resources already deemed to be of relevance or importance to other libraries, thus, effectively performing the function of a peer-review. Whilst each database is significantly smaller than web, there is, proportionally, a greater degree of quality resources present in each. The first port of call for the selector can then be the smaller, peer-selected database.

Conclusions

What we have seen, then, is that both INTERCAT and CORC have had some degree of success. This success has differed for each, with INTERCAT serving to enhance existing services, whilst CORC has become an entirely new service. Inherent in both is a bedrock of co-operation, again slightly different in scope. For INTERCAT co-operation was on the already existing, well established lines of the bibliographic network; Inter-library co-operation as resource sharing. CORC, continues this approach, but, by virtue of no longer being tied to MARC-only metadata, expands co-operation to the cross-sectoral or cross-domain level. In addition to this, a practical implementation of CORC within the library challenges the functional model of library organisation and engenders intra-library co-operation.

Virtual collections within the hybrid library will continue to challenge for as long as we seek to bring them under bibliographic control. The ideal of inter-working that grounds our cross-domain concerns is the realisation of co-operation. Neither INTERCAT nor CORC can take the credit for this achievement, but they can and do hold a mirror to library practices and serve as model, either historic (as in the case of INTERCAT) or current (as with CORC) of how they hybrid library functions within and beyond itself in meeting the challenges of resource discovery on the web.

ⁱ INTERCAT project, <http://www.oclc.org/research/projects/intercat.shtm>

ⁱⁱ *ibid.*

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

^{iv} Caplan's use of Stueart and Moran (1993) in "Redesign of Library Workflows: Experimental Models for Electronic Resource Description", <http://lcweb.loc.gov/catdir/bibcontrol/calhoun.html>

^v *ibid*

^{vi} This has been noted by Clifford Lynch: "The New Context for Bibliographic Control In the New Millennium", <http://lcweb.loc.gov/catdir/bibcontrol/lynch.html>