

CORC at Cornell Project: Final Report

By the CORC Team:

Karen Calhoun (coordinator)
Martha Hsu
Yumin Jiang
Jill Powell
Don Schnedeker
Pam Stansbury
Bill Walters

Contents:

[Executive Summary](#)

[Introduction](#)

[Workflow](#)

[Highlights and Issues of the Project](#)

[Selection and Acquisitions Information](#)

[CORC's Potential for Public Services](#)

[Comparison of Dublin Core and MARC Records](#)

[Assessment of the CORC Database](#)

[Evaluation of CORC: Recommendations for OCLC](#)

[Recommendations for CUL](#)

[Appendix 1](#)

[Appendix 2](#)

Executive Summary

CORC—the Cooperative Online Research Cataloging project—is a collaborative research initiative of the OCLC Office of Research and about 150 participating institutions, including Cornell. It provides Web-accessible shared databases and automated tools to help libraries manage and provide intellectual access to the massive amount of material becoming available on the Web. The "CORC at Cornell" project was undertaken by a small, cross-functional research team of seven people. We worked on the project from mid-May to mid-November 1999. In the approximately 400 hours we spent on the project, we:

- selected and cataloged about 120 new resources for the library catalog and Gateway
- experimented with a new workflow in which selectors, reference staff, and catalogers collaborated to create metadata
- gained practical experience with Dublin Core
- explored the features of CORC for creating subject guides (a.k.a. *pathfinders*)
- evaluated the functions and features of the CORC system
- enhanced our abilities to select, manage and control electronic resources for the library

In the course of the project, we developed a [project Web site](#) to share information about our activities inside and outside Cornell. We gave a number of interim reports to library committees; in addition, there was a good deal of

interest outside Cornell in our cross-functional approach to the project.

Gaining insight into new ways in which selection, description and access can work together was a highlight of the project for the Cornell team. Particularly valuable aspects of CORC are the interoperability of Dublin Core and MARC and the system's support for having selectors and reference staff participate in the resource description process. The body of this report elaborates on the numerous issues that our team discussed in the project, such as:

- Who should do selection of networked resources? How can the library take greater advantage of what public services staff know about Internet resources?
- Who should create metadata for our networked resources, and what do they need to know and be able to do?
- How much is enough in a record for a networked resource?
- In addition to descriptive metadata, selectors and acquisitions staff need evaluative and managerial metadata. How might Cornell build the systematic collection and online storage and retrieval of this kind of metadata into the workflow?

Our experiences in the CORC at Cornell project have led to a number of recommendations, which are briefly summarized below. We anticipate discussing this report and our recommendations with our colleagues at the February 2000 Academic Assembly meeting. For details, refer to the full report, which follows.

1. Cornell library staff should have the option to use CORC to precatalog networked resources, using Dublin Core, and for other purposes.
2. Cornell library staff should prepare brief guidelines for the use of Dublin Core in our library.
3. Further research should be undertaken to define the contents of a "useful" record for a networked resource, from a reader's perspective.
4. CORC training sessions should become a regular part of the CUL staff training program.
5. Reference librarians should have a broader role in the selection of networked resources.
6. Cornell library staff should explore means for building the systematic collection and online storage of evaluative and managerial metadata into the e-resource workflow.

[Back to Top](#)

Introduction

This report is intended for the Cornell University Library (CUL) administration, for the OCLC Office of Research, and for others interested in options for networked resource selection and cataloging at Cornell.

[CUL's tactical plan](#) calls for integrating traditional and digital resources and services in support of the Cornell community's information needs. It also gives priority to service improvements, collection building, and organizing the collection for effective use, including the following objectives:

- Continue and accelerate the development of digital resources and services, where content is fully integrated with bibliographic finding tools ...
- Expand access to electronic resources in a variety of formats ...
- Insure access to electronic resources through metadata, cataloging, and other means
- Increase cooperative cataloging programs with other institutions ...

OCLC's Cooperative Online Resource Cataloging (CORC) research project has the potential to help CUL's librarians move more quickly toward managing the massive amount of material becoming available on the World Wide Web and toward identifying and providing intellectual access to relevant, high quality information there. In addition, CORC is based on a collaboratively built, shared databases and cooperative creation of both metadata and finding tools.

The OCLC Office of Research runs CORC, and as of this writing about 150 institutions are participating in the project. The CORC experimental system puts in place a framework for cooperative cataloging of Web resources that features

mapping between Dublin Core and MARC; automated tools for the creation of metadata records in Dublin Core and MARC; export of metadata in HTML, RDF and MARC; import, generation, sharing and export of subject guides; and numerous other features. For more information visit <http://www.oclc.org/oclc/research/projects/corc>.

The library became a CORC participant in May 1999. We began with training sessions conducted by OCLC Office of Research staff (in CORC) and Diane Hillmann (in the use of simple Dublin Core).

Our team structured participation in CORC as a research project for the purpose of:

- Selecting and cataloging new Internet resources of interest to the CUL community
- Experimenting with a new workflow, in which a reference librarian or selector generates a preliminary record in Dublin Core and the cataloger completes it in MARC format
- Experimenting with the centralized, cooperative creation of subject guides (called *pathfinders* in CORC)
- Gaining a greater understanding of and practical experience with Dublin Core and automated record creation tools
- Providing free access to the CORC database across campus for searching

We pursued the first goal in several ways. Selectors independently identified a number of resources to be processed via CORC; in addition, we set up mini-projects to catalog selected resources from ICE (Internet Connections for Engineering) and the BII (Business Internet Index), which are sets of Web pages developed by CU's Engineering and Management libraries, respectively.

Table 1 gives the names of research team members and their roles. We estimate the team devoted a total of 400 hours to CORC-related work from mid-May to the end of the project (includes team meetings, e-mail communications, work on the project Web site, time spent in training, meeting and presentation preparations, reading and study, and CORC online usage). In that time we produced about 120 new records for the NOTIS and Gateway catalogs and about half a dozen pathfinders.

Table 1. Project Staffing

<i>Role in Project</i>	<i>Name</i>
Project coordinator	Karen Calhoun
Original cataloger	Yumin Jiang
Copy cataloger	Pam Stansbury
Reference librarian	Jill Powell
Unit library director	Don Schnedeker
Collection development librarian	Martha Hsu
Collection development librarian	Bill Walters

In the course of the project, we developed a [project Web site](#) to facilitate the dissemination of information about Cornell's project inside and outside Cornell. The site includes a Web-accessible Powerpoint presentation describing and demonstrating CORC, links to team meeting notes, a link to our initial proposal to the library administration, a link to the Dublin Core guide edited by Diane Hillmann, a link to CORC itself, and other links. Two members of our team

attended CORC Participants' Meetings in Dublin Ohio (one in April, one in November). Here at home, we gave interim reports and solicited feedback at meetings with Ross Atkinson and Janet McCue (who served as our project sponsors), the Technical Services Executive Group, the Public Services Forum, General Selectors, and the Instruction and Reference Policy Committee. In February 2000 the team anticipates discussing this report and its recommendations with members of Academic Assembly.

There was a good deal of interest outside Cornell in the cross-functional approach we took to CORC participation, and the project coordinator was invited to present the details at an ALA Annual meeting on the topic of CORC. In addition, Cornell's CORC project is featured in the 1999 OCLC annual report.. Finally, partly as a result of interest in Cornell's project, the project coordinator was offered the opportunity to co-edit a special issue of the *Journal of Internet Cataloging (JIC)*, which is to be devoted to research papers on CORC. Publication is scheduled for next summer.

[Back to Top](#)

Workflow

The *existing* production workflow for cataloging networked resources generally includes the following steps:

1. A selector identifies and selects an e-resource
2. The selector initiates a request to acquire and/or catalog the resource (generally using the "Networked Resource Selection Form" at <http://www.library.cornell.edu/tsmanual/EdocsFORM.html>)
3. Acquisitions and selector exchange inquiries as needed
4. Acquisitions negotiates with vendor/publisher/author (for licensed resources)
5. Acquisitions prepares preliminary data for resource description (in MARC format) and forwards the selection form to cataloging
6. Acquisitions/catalogers/selectors exchange inquiries as needed
7. Catalogers consult e-resource, cataloging standards and databases to revise and complete preliminary data for resource description
8. Catalogers/acquisitions/selectors/information tech. staff exchange inquiries as needed
9. Cataloger produces resource description for NOTIS catalog and Gateway

Everyone is aware of the millions of Internet resources and the astonishing growth rate of the Web. The authors of *Getting Mileage Out of Metadata* (Chicago: ALA Publications, 1999) put it nicely: "The central paradox of the Web has proven to be that the more information available on a subject, the greater the likelihood that relevant, authoritative information will not be found." As a result, never has metadata—used to support the discovery, retrieval, accessibility and control of resources—been more important. We wanted to take advantage of the opportunity provided by the CORC research project to delve into the following questions, which are being asked in libraries everywhere:

- Can the existing selection, acquisitions and cataloging system be modified to make resource descriptions (metadata) available more quickly? In greater numbers? At less cost?
- If we changed the existing system for producing resource descriptions, would it deliver records that are useful to readers?

To gather insights into these questions, the CORC team developed an *experimental* workflow to be used by team members. The new workflow includes steps similar to those in the existing workflow except that (a) selectors prepare the preliminary records, using the Dublin Core standard, and (b) reference librarians as well as selectors identify, choose, and create preliminary records for Internet resources. Later, catalogers finish the records in MARC format and export the metadata to the NOTIS catalog and thence to the Gateway. In our research project, to keep the scope of the project manageable, we dealt with unlicensed Web resources only.

The Dublin Core (DC) elements, with their readily understood names and labels, were designed for flexibility and ease of use by those without traditional cataloging expertise (for more information, visit <http://purl.oclc.org/docs/core/index>).

[htm](#)). Indeed, the DC's initial purpose was author-generated description of Web resources capable of producing consistent, helpful indexing. Today, the DC set of elements stands poised to become the preeminent standard for descriptive metadata of Web resources. In CORC, the DC elements are mapped to MARC record elements for conversion and export. The mapping also supports the CORC database, which is an interoperable one of both DC and MARC records.

Our experience in the CORC project suggests that the changes we tested to the workflow can ease and streamline the production of Web resource descriptions. It also suggests that selectors and reference staff can readily use DC to create preliminary records using CORC. As will be discussed later, we have little data for speculating about the utility of records, from a reader's perspective, but it is possible that the records we are producing are more than adequate, and may even go beyond what CUL readers find essential.

[Back to Top](#)

Highlights and Issues of the Project for Team Members

The following list, drawn from team meeting notes and e-mails, summarizes team members' remarks on the highlights and issues of the CORC at Cornell project.

Highlights

- Doing things a new way
- Seeing how selection, description and access work together (the whole workflow, not just one part of it)
- Learning about and using Dublin Core
- Not having to use the Networked Resource Selection Form; searching the CORC database to find vetted resources or to see descriptions of resources, and transferring the information into a Dublin Core template for use locally
- Having selectors and reference staff begin records that can be built upon in cataloging
- Interoperability (being able to flip back and forth to view records in DC and MARC)
- Responsiveness of OCLC Office of Research staff to questions and suggestions
- The ability to share our records and use others' records in an international shared database for networked resources
- Finally understanding how to use the CORC pathfinder functionality (esp. for those with less sophisticated HTML skills)
- Finding pathfinders with resources grouped together (an advantage of CORC for selectors)

Issues

- Who should do selection of networked resources? How can the library best take advantage of what public services staff know about Internet resources, and use that knowledge in the selection process? Many public services staff do some selection now, but should the opportunity to select be opened to everyone? Selectors have an important role to play in vetting selections and maintaining control of what is selected. How can we achieve the best of both worlds? Since selectors are very busy with licensed resources, would it make any sense to invite reference staff to take on a larger role in selecting nonlicensed resources?
- Selection criteria for the CORC database. In discussions inside and outside the team, some bibliographers expressed concern about the absence of collection development standards for the CORC database. Team members did not agree on this point; some pointed out that CUL itself has no documented, system-wide criteria for selecting for the Gateway, so it may be unrealistic to expect the heterogeneous contributors to the CORC database to reach consensus on a uniform collection development policy.
- How can we effectively store, retrieve, and share evaluative and managerial (e.g., licensing, troubleshooting) metadata of importance to selectors and acquisitions staff within CUL? Neither MARC nor DC records are adequate for storing this kind of information. How might CUL build the systematic collection and online storage of this kind of metadata into the CUL workflow for networked resources? What sorts of information should be stored? This issue is discussed further later in this report.

- In order for selecting materials on CORC to work, access to the CORC database and the online template must be quick and convenient. Adjusting subject headings and fields in the record must not be cumbersome. The current CORC experimental system needs work in this regard, as will be discussed later in this report.
- Who should create metadata for our networked resources? CORC enables us to spread the wealth (of resource description) outside the technical services units. Based on our experiences in the team, it does appear the average selector or public services librarian could deal with using CORC to create preliminary records to be completed by catalogers. The skills required to use CORC are basic familiarity with computers and the ability to move around in, and manipulate, online forms and records. With a little training, this should not be problematic for any library staff member. Without having to be a cataloging specialist, the staff member has the opportunity to create an online textual description of the resource, with key words to make it findable on the Gateway. The cataloger uses the information as it was created by the staff member and converts it to MARC, improving efficiency and accuracy. Many staff members are dissatisfied with the paper-based selection form we are using now to pass along selection information to acquisitions and cataloging, and having technical services staff start from scratch with each Internet resource description is wasteful. But if selectors and reference staff begin creating preliminary records, how much would be expected of them, in terms of record content? Should students and acquisitions staff be taught to use CORC and DC to obtain preliminary records?
- Related to the issue above—we are not sure how to implement the Dublin Core element set. Should there be guidelines? Would it make sense to agree on some basic guidelines for the content of a CUL DC record from CORC (like the [University of Minnesota library](#) has done)? We are sure we don't want using DC to be tedious, time-consuming or complicated, so if we have guidelines, they must be simple and straightforward to teach and use.
- How much is enough in a record for a networked resource, when the object itself is one click away from the metadata? We found ourselves musing over this issue quite a bit. One discussion focused, for example, on assigning full LC call numbers to networked resources. Some team members thought this practice should stop, remarking that public services staff often don't use the call numbers for anything, and some feel a call number's presence on a record for a networked resource is actually confusing. There is little searching by patrons or staff by call number; some suggested that the time catalogers now spend assigning call numbers could be more usefully spent doing something else. On the other hand, a recent public discussion on the CORC-L list revealed staunch supporters of assigning call numbers to everything.
- In discussing CORC pathfinders, team members were drawn into considering the broader question of the role of subject guides in CUL. Also, the concept of a *shared, multi-institutional* database of pathfinders is a new one for public services librarians. Do these issues make any difference in whether and how we might make use of the CORC pathfinder functionality within CUL?
- CORC performs best with a high performance PC (300 MHz or better). Not all members of the team, and not all staff members who might make use of CORC have high performance PCs.

[Back to Top](#)

Selection and Acquisitions Information

As already mentioned, CORC can allow collection development and acquisitions staff to assume a broader role in the CUL networked resources workflow. Selectors can begin the record-creation process in Dublin Core format, entering basic descriptive information. Of particular importance, they could also enter those fields that rely on special knowledge of the resource or the acquisitions process. For example, selectors might enter fund codes or evaluative notes; acquisitions staff might enter license restrictions, access information, or information for trouble-shooting.

The full potential of the workflow with which we have experimented can be realized only if selectors and acquisitions staff can use a database—CORC or an alternative—that includes those fields and features most useful to them. If selectors want to include notes that discuss the relationships among two or more resources, for example, then the database should include a field where this information can be provided. If acquisitions staff want to indicate whether a license allows for the

fulfillment of ILL requests from other institutions, then the database should provide a place for this information.

[Appendix 1](#) describes the fields and features identified as potentially useful by collection development and acquisitions staff at Cornell University. The list is the result of work coordinated by Bill Walters, in collaboration with Scott Wicks and David Block. Specifically, these are the items of information required or desirable in a generic database of Web resources for use by library staff. (We assume that much of this information would be available to the public as well.) While the list includes many items that are applicable to CORC, it is intended to be more broadly useful. That is, the fields that might be useful in CORC are often the same fields we'd like to see in a library management system, online catalog, or resource database linked to catalog records. Some of the fields in our list have equivalents in MARC, DC or LMS records, or some combination thereof; some do not. The next steps of this analysis, which are not provided here, would be to identify those fields that are required or highly desirable, but that are poorly accommodated in a MARC or DC database and/or in our new local system, Voyager.

The list is presented in three categories: (1) what we believe are required fields and features; (2) optional fields and features that should be included; and (3) optional fields that are less important but worthy of further discussion. We have omitted the basic fields that are already in widespread use (title, author, call number, subject headings, check-in information for serials holdings, etc.).

[Back to Top](#)

CORC's Potential for Public Services

Reference librarians have long compiled bookmarks and web sites of interest to their patrons. Staff at Cornell's Engineering Library began a series of web pages in May 1994 called [ICE](#) (Internet Connections for Engineering) that arrange web sites by subject (aerospace engineering, electrical engineering, etc.).

The ICE pages are very labor-intensive to maintain. Resources need to be double-listed under several subjects and the links frequently go out of date. The Engineering Library has no usable searchable index to these sites. As these sources get cataloged on the Library Gateway via CORC, we can rely more on the Gateway instead of ICE. This will be less labor intensive for Engineering Library staff and allow sharing of the resources with a broader audience. They will be accessible from a central location (the Gateway) where they are indexed and searchable.

The CORC team has been working through the alphabet of ICE, and many now have complete cataloging. Although we are bringing the CORC research project to a close, Engineering Library staff would like to continue identifying and contributing sites to CORC for inclusion in the Gateway. They estimate two to four hundred entries remain.

CORC provides a convenient way for reference librarians to select and precatalog electronic resources. Engineering Library staff report that the lack of paperwork involved is wonderfully liberating, that browsing through other institutions' electronic records is very insightful, and that being able to help patrons by having the sources readily accessible is fulfilling our library's service mission.

We believe our project—in particular our progress with ICE—demonstrates that it is both beneficial and practical for reference staff to take a broader role in the selection and description of networked resources. Since CUL selectors are so busy with licensed resources, it may be especially appealing to leverage the knowledge, experience and skill of CUL reference staff to add nonlicensed Web resources to the collection.

Staff at the Engineering Library have developed Engineering Research Guides, also by subject, which point to cataloged resources, most of which are subscription-specific (indexes, e-journals) as well as pointing to ICE. These could be developed as pathfinders; however, we feel the CORC pathfinder software needs more work before someone proficient in HTML coding would gain from altering his or her current practice of creating subject guides and instructional materials directly in HTML. On the other hand, authors of subject guides and instructional materials wouldn't have to learn HTML in order to create these documents in the CORC pathfinder database. Even so, the CORC system is slow for

older computers (given our experience with the Power Mac 7300/180) and doesn't always import HTML files accurately. When a CORC team member developed two pathfinders ("Patent Resources on the Web" and "Engineering Research Guide for Computer Science"), CORC didn't integrate the introductory paragraph but wanted only to create HTML links with it. Pathfinders aren't just a list of URLs with descriptions; they could have introductions, tables, and more. Sharing the pathfinders with other institutions would be terrific, so we encourage further development in the pathfinder software so that we could possibly use this aspect of CORC in the future.

[Back to Top](#)

Comparison of Dublin Core and MARC Records in CORC

The original MARC format was created in mid 1960s, and since then the revision and expansion of the MARC formats has been an ongoing process. A MARC record involves three elements: the record structure, which conforms to national or international information exchange standards; the content designation, which are the codes and conventions identifying data elements in a record; and the content, which is mostly defined by other standards such as Anglo-American Cataloging Rules. Formats are defined for five types of data: bibliographic, holdings, authority, classification, and community information. Library of Congress maintains the USMARC format in consultation with various user communities.

The Dublin Core (DC) standard is a simple element set intended to facilitate discovery of electronic resources. It contains fifteen basic elements (see Table 2). It was first developed in 1995. Although describing electronic resources is the goal, the DC standard is flexible enough to allow description of other kinds of resources. Since its inception, it has gained wide recognition and/or acceptance in the resource description communities such as libraries, museums, and government agencies. Some important characteristics of the DC standard include: simplicity, semantics interoperability, international scope, and extensibility.

The basic elements are designed to serve as a lowest common denominator kind of description standard, and each element has a very broad definition, sometimes to the extent of vagueness. To address this issue, there are ongoing discussions and efforts in the DC community in further extending beyond the fifteen basic elements (also known as unqualified DC). Various qualifiers under each element have been proposed, and the CORC system incorporates many of the qualifiers suggested by DC working groups. In addition, the OCLC CORC staff has also recommended some new qualifiers.

As the DC standard continues to develop, the crosswalk (mapping) between MARC bibliographic data elements and the DC is also evolving. CORC participants have the choice of using either the MARC or the DC standard, and CORC features a crosswalk between the two to enable easy viewing and inputting of information in either a MARC record or its DC counterpart. The OCLC CORC staff has collaborated with Library of Congress, which has just released [updated DC/MARC crosswalk](#) recommendations.

The crosswalk outlined below in Table 2 is primarily based on Cornell CORC team's understanding of the DC elements during the length of our participation. It includes the following information:

- Commonly used MARC fields for describing electronic resources in full-level records
- Fifteen basic elements of DC
- Cornell University Library OPAC labels
- PCC core-level record requirements for monographic computer files, and CONSER core-level record requirements for serials

Table 2. MARC and DC Parallels in CORC

Codes: M – mandatory, MA – mandatory if applicable, O – optional, N/A – not applicable

MARC Tag	Dublin Core	CUL OPAC Label	PCC Requirements
----------	-------------	----------------	------------------

100 or 110	Author or Creator	Author/Name	MA
245	Title	Title	M
260	Publisher	Published	M
260	Date	Published	M
310	Description?	Frequency	O (CONSER)
362	Description?	Published	MA (CONSER)
500 (General note)	Coverage	Notes	O
500 (General note)	Description	Notes	O
500 (Source of title)	Description	Notes	MA
506 (Access restriction note)	Description	Notes	O
516 (Type of computer file)	Type	Notes	O
520	Description	Summary	MA
538 (Mode of access)	Format	Notes	M
540	Right	Notes	O
546	Language	Notes	O
6xx	Subject	Subjects	MA
700 or 710	Contributor	Other Names	MA
856	Identifier	Electronic Access	MA
787 (Nonspecific relationship entry)	Relation	N/A	O
786 (Data source entry)	Source	N/A	O

Current cataloging policy at Cornell calls for cataloging electronic resources at full level. PCC and CONSER have also formulated core-level record standards which require fewer data elements (mostly notes fields). Important data elements in PCC core-level records or CONSER core-level records include author, title, publisher information, subject headings, and Internet address (URL). The DC standard, on the other hand, has no required elements; every element is

optional. Nonetheless, most DC records, especially DC records in the CORC database, have at least some elements. As noted earlier in this report, we have pondered the questions How useful are all the different information elements to our library users? What level of cataloging is more meaningful for users? What kind of cataloging standard is appropriate for electronic resources? The next section of this report will discuss this issue in more detail.

The crosswalk is a critical CORC feature for our project team as it enables library staff other than catalogers to participate in resource description, especially in creating preliminary records using the DC standard. However, the automatic matching between the two records can sometimes problems. As several CORC participants commented in their email messages sent to CORC-L, MARC and DC are different standards, and information in one record may not translate very well to the other. As the table shows, some fields in MARC have no good equivalent in DC, and some DC fields map to MARC fields seldom used in ordinary MARC records.

The MARC standard facilitates cataloging of many kinds of materials. For example, the format for serials includes information elements such as frequency of publication, the numeric and/or chronological designation of the first issue, etc. The DC standard is less well developed in this regard. During the project, we have cataloged many different types of Web resources. Catalogers on the team feel that the basic DC elements are adequate and efficient for describing many of the general Web sites, but not quite as useful for resources such as electronic serials and resources that have complex relationships with other resources. Some important information about a serial (e.g., frequency, beginning issue) and complex relationships (e.g., preceding and succeeding titles, supplements, different physical manifestations) does not fit well into any basic DC elements.

[Back to Top](#)

Assessment of the CORC Database

The CORC project has sparked many lively conversations about CORC records both within our team and outside Cornell. Everyone seems to be trying to define what and how much to include in a record for a networked resource. As near as we can tell, the answer is "it depends." The definition of a "useful" record is intertwined with the role, experience, and expectations of each person who is engaged in considering the issue, and what is deemed "useful" is very much in the eye of the beholder.

There does seem to be some common ground. The presence of even skeletal DC metadata enables better indexing and retrieval of Web resources than what users have come to expect when they use the current generation of commercial search services.

As has already been noted, all of the elements of DC records are optional, repeatable, and displayable in any order desired. The standard has tremendous potential for libraries. DC metadata can be embedded in Web documents in HTML <META> tags or in [RDF/XML syntax](#). DC metadata can also exist as independent, freestanding records, which can be collected in resource files—like the CORC database. Fully implemented at Cornell, DC metadata could support faculty-generated indexing to e-reserve material or faculty/departmental Web pages; could support the indexing of Cornell dissertations; could be taught to students in library-sponsored HTML classes, so that they become more effective Web authors; might coexist in the same catalog with our MARC records; might be mapped to other metadata standards used in digital library projects; could be used to support searching of library-created Web pages (as has been done for the new Gateway Help); and other uses.

DC, however, moves beyond the reliable and well understood construct of the MARC record, with its ties to AACR2 and a century of cataloging tradition, to a more flexible concept that is being used to encode just about anything related to an Internet resource. DC records can range from the lengthy and content-rich to the skeletal. DC elements are understood and used differently by those creating DC records. There is currently no consensus about DC content standards—even about "best practices"—and this "anything goes" atmosphere, coupled with the fact that CORC is a research initiative rather than a production system, has led to a widely heterogeneous mixture of practices represented in CORC database records.

Viewed from the perspective of library staff members who expect that useful records conform to known, consistent standards and contain the fullest information possible, DC CORC records are frequently not useful. In a recent article, Norm Medeiros notes "to librarians, the thought of abandoning [MARC], which has millions of records already invested, is heresy..." (*Online* (Nov/Dec 1999): 57-60)

Is it possible to reconcile the present "Wild West," freewheeling atmosphere of CORC record creation with the standards-based precision of MARC and AACR2 and the expectations of library staff members? We are not sure, but we believe that the attempt should begin by understanding what the *readers* who must use the metadata we produce want, expect and need.

Lundgren and Simpson, working at the University of Florida, studied reader preference for particular metadata elements and published their results this year (*Journal of Internet Resources* 1 (4: 1999): 31-44). In their article they review the few studies that have yielded user feedback on the cataloging of library materials and note that historically, decisions on the content of catalog records have *not* been based on user-defined preferences. Their study attempts to provide a framework for determining the most-desired elements in an Internet resource description, based on feedback drawn from survey responses from about 450 randomly-selected University of Florida graduate students. The most highly rated elements were title, primary author or creator, Internet address (URL), and summary note or abstract (DC. Description.Summary or MARC 520 field). Students expressed agreement regarding the value of other record elements, but these were the four key elements.

In the CORC project we gained Lundgren and Simpson's permission to use their survey instrument in our own study. However, in a pretest of the survey instrument with members of several Cornell unit library advisory boards, only 6 of the 26 survey pretests were returned. While there are too few responses to draw any conclusions, the survey instrument does appear to be acceptable for gathering the information, and the 6 responses we received are consistent with the findings of the Florida study. Even across 6 responses, a strong pattern is evident in what is regarded as most important in a description of a networked resource--title, summary note or abstract, primary creator, and Internet address. We feel that further efforts to carry out the survey at Cornell are justified and would yield worthwhile information to guide future library decisions about metadata content. Along the same lines, we feel that the most productive conversations about DC and MARC will assume that both have their place at Cornell. We should focus on how to forge a complementary relationship between the two standards, striving to optimize the strengths of each.

[Back to Top](#)

Evaluation of CORC: Recommendations for OCLC

CORC includes a number of tools to support the creation and editing of metadata records in Dublin Core and MARC. We applaud the groundbreaking work that the Office of Research is doing in this area. While the tools are generally helpful, and OCLC Office of Research staff have been extremely responsive and consistent in improving them over the course of the project, many need further enhancement. We also believe that in the production CORC system, OCLC should encourage conformance to some very basic, minimal content guidelines for DC records, keeping in mind the great importance of keeping DC easy and straightforward for those not trained as catalogers. [Appendix 2](#) offers our comments and recommendations for OCLC pertaining to specific editing tools and to pathfinders.

Recommendations for CUL

Many library staff members have their hands full with the Endeavor implementation. Bringing up Voyager next summer is clearly a higher priority for the library than figuring out the next steps for CORC. For that reason, we have abbreviated our project, which was originally scheduled to run through the end of 1999. For the same reason, and because OCLC has announced it is moving CORC into production in summer 2000, we believe that implementation of the following recommendations should begin no sooner than next summer.

1. CORC should be available as an option for Cornell staff to use for precataloging of networked resources, using Dublin Core. CORC could also be used to generate DC metadata for our own Web pages (e.g., Gateway Help was prepared using

DC metadata, though it was not done using CORC).

Who: acquisitions/students/selectors/reference staff

When: following training (see below)

2. Basic brief guidelines for the creation of DC precataloging records at CUL should be prepared, following the [University of Minnesota](#) model. Required fields should be title, URL, a summary/annotation, and a note indicating who selected the resource. (The last note is temporary; its purpose is to give the cataloger a contact name, in case he/she has questions.)

Who: A member of the CORC team, in cooperation with the editor of the Gateway cataloging guidelines

When: before training (see below)

3. CUL staff should have the option to search the CORC catalog and/or pathfinder database as a resource file; as a backup to the CUL gateway and search engines; as a source of summary notes for catalog records; as an aid to selection; for help in creating or for cloning a subject guide or for sharing a subject guide with other institutions; and for other uses.

Who: all CUL staff with an interest

When: following training (see below)

4. CORC training sessions should become a regular part of the CUL staff training program.

Who: Selected members of the CORC at Cornell team, with assistance from LHR; and/or Nylink trainers (if available and appropriate).

When: After CUL DC guidelines and OCLC CORC documentation/help are ready.

5. Reference librarians should have a broader role in the selection of nonlicensed networked resources.

Who: Reference librarians

When: Following discussion and action by LMT and other appropriate groups.

6. Continue to collaborate on cataloging selected ICE titles for the CUL catalog and Gateway. Undertake similar collaborate projects (e.g., for the BII) as requested, depending on timing and availability of necessary staff resources.

Who: Engineering Library staff and catalogers in CTS; others as willing and able.

When: Ongoing, beginning now.

7. CUL should develop a database to store internal evaluative and managerial information related to networked resources (see Appendix 1 for a description of what the database might contain). A small work group should be appointed to explore if and how information in this database could be linked to Voyager records for these resources.

Who: Work group appointed by LMT and appropriate IT staff

When: Following discussion and action by LMT

8. Further research should be done to define the contents of a "useful" record for a networked resource, from a CUL

reader's perspective.

Who: To be determined

When: To be determined

9. As opportunities arise to move in the direction of an interoperable database for Voyager (i.e., one that would allow the coexistence of multiple metadata formats), CUL should pursue them.

Who: To be determined.

When: To be determined.

[Back to Top](#)

Appendix 1: Data Fields to Support Selection and Acquisitions Activities

This appendix describes the fields and features identified as potentially useful by collection development and acquisitions staff at Cornell University. The list is the result of work coordinated by Bill Walters, in collaboration with Scott Wicks and David Block. Specifically, these are the items of information required or desirable in a generic database of Web resources for use by library staff. (We assume that much of this information would be available to the public as well.) While the list includes many items that are applicable to CORC, it is intended to be more broadly useful. That is, the fields that might be useful in CORC are often the same fields we'd like to see in a library management system, online catalog, or resource database linked to catalog records

(1) Required Fields and Features

Selector — Person responsible for evaluating the resource and determining its suitability for the collection.

Type of publication — Monograph, serial, etc. If serial, indicate whether it is updated continuously (as many databases are) or in discrete parts (issues, editions, etc.). Do new parts cumulate or supersede previous parts?

Genre — Full text, bibliographic, numeric, etc. Leave space for multiple genres.

Access — Is the resource or aggregation freely accessible to all users? To the entire university community (including visiting scholars, emeritus faculty, alumni, distance learners, etc.) or only to currently registered students, active faculty, and currently employed staff? To only those students and staff affiliated with certain departments or programs? Only from certain IP addresses? (List address ranges.) Only from on-site locations (individual libraries, buildings, etc.)? Is access available to walk-in users not affiliated with the university?

Number of simultaneous users — Mention any special provisions for temporary use by larger numbers of workshop or seminar participants.

Price — Include special pricing information. Discount if ordered by a particular date? Discount based on membership in consortia or other organizations? (Include membership number.) Discount tied to continuation of print subscription? (Include link to record for print version, if applicable.) Are there separate costs for content, print access, and electronic access? Or are print and electronic versions priced separately?

Payment history — Payments made, payments due, payments anticipated in future.

(2) Optional Fields and Features

Description for public catalog — Summarize purpose, content, and intended audience. This field is strongly recommended for aggregations, reference resources, and data files.

Library's representative — Person responsible for negotiating with the publisher and initiating subsequent inquiries. Usually the acquisitions librarian, although this role may be delegated to someone else (a selector, government information librarian, etc.).

Publisher's representative — Name, address, e-mail address, and phone number of the individual representing the publisher or distributor. Indicate when contact information was last updated.

Fund code.

Subject or department code (a separate field — not part of the fund code).

Endowment budget code (a separate field — not part of the fund code).

License highlights — Does the license allow for printing? Downloading? Sharing with outside users (interlibrary loan)? File transfer in electronic form? Transfer in printed form? Educational use only? Duration of the license? Who keeps back issues if the license expires? Do we own or lease the resource? (In general, the license fields should indicate which uses are authorized — not which uses are not authorized.)

Link to full-text image of the complete license agreement.

Access/verification mechanism(s) — IP verification? Proxy server? Password? (Indicate password; not for public display.)

Requester (patron requesting the item) — Include contact information along with any special concerns.

Special hardware requirements — Specify the hardware and its availability (in the Reference room, elsewhere in the library, in the computer labs, etc.).

Special software requirements — Include both operating systems and proprietary reader/viewer/analysis software. Indicate availability (free on the web, networked in accordance with site license, available at particular locations in accordance with site license, distributed by special arrangement, etc.). Provide link to software information, if available.

Edition — A newer or older edition of another work? Reprinted? Updated? Supersedes earlier work? Does not supersede earlier work?

Format — Identical in content to a print resource (or another web resource)? Identical in appearance (layout, pagination, etc.)? (PDF images of a print resource are usually identical to the print copy in both content and design. Most other formats — HTML, etc. — may be identical in content but are not identical in appearance.)

Component of an aggregation or series? — List aggregations and super-aggregations of which the resource (or aggregation) is a discrete component. Provide links to aggregations of which it is a part. Indicate whether these aggregations have been cataloged separately.

Includes constituent components? — Is the title an aggregation (or super-aggregation)? What components does it include? Provide links to component parts. Indicate whether these parts have been cataloged separately.

Use statistics — Are use statistics available? How (via web? via e-mail? via printed report?) Who compiles these

statistics, either within the library or elsewhere? (Include contact information.)

Temporary notes — Include one or more fields for internal notes sent among the library staff (requests for action, notes on dealing with publishers' representatives, etc.). These notes should be temporary. That is, they should be automatically removed from the record after a specified number of days, or when a specified action has been taken.

(3) Additional Fields to Consider

Links to files of e-mail correspondence with publisher or distributor. (Create a file in which to store all correspondence concerning the title or aggregation.)

Basis for collection development decision — Flyer or announcement, review in journal or selection guide, recommendation from patron (include contact information), trial subscription, examination of the resource itself, etc. Allow room to choose more than one option.

Selection status — Waiting for trial access; waiting for additional information; sent to liaison; sent to another library within the university; sent elsewhere (outside the university); decision pending; decision made (yes, no, or reconsider at a later date). Provide details for each status (when was trial access requested? waiting for what information? sent to which departmental liaison? what was liaison's recommendation? when was final decision made?).

Reason for "no" decision — Out of scope; high absolute cost (added to desiderata file?); high relative cost; better alternatives available; low quality; bias in content or presentation; cannot be supported adequately by library staff; etc.

Descriptive and evaluative information for library staff — This field supplements the description in the public catalog. Assessment of author; assessment of publisher (if unusual publisher); assessment of authoritativeness; intended audience; level of presentation; similar resources available elsewhere?; comparison with other resources; inclusiveness (comprehensive or selective); breadth of coverage; depth of coverage; likely to be supported in future by publisher/distributor?

Are holdings complete? (for web equivalents of print resources) — Yes; recent only; archival only; other.

Currency — Is updating done regularly? Does web version appear before or after the print edition?

Preservation plans and preservation history — Scheduled format migration; scheduled conservation (for web resources also held on physical media); record of preservation activities. Backup copies held by the library? Special considerations (irreplaceable, held by few libraries, etc.).

Deselection — When? Why? By whom? Disposition of physical copies, if any.

[Back to "Selection and Acquisitions" Section](#)

[Back to Top](#)

Appendix 2: Evaluation of CORC Record Creation and Editing Functions: Recommendations for OCLC

The tools developed by OCLC to support automated cataloging are generally helpful. We applaud the groundbreaking work that is being done by the Office of Research and the responsiveness of OCLC staff in working with us on CORC. User support, system support, and turnaround time for enhancements have been outstanding. We offer the following comments and recommendations for OCLC. They pertain to specific editing/record creation tools, CORC database records, and pathfinders.

Harvester. In general the harvester is a great time saver. It does a good job of capturing basic information from the HTML code of the Web site. By itself, the capture of a URL, so that it does not have to be keyed (possibly with typos) by the user, is a great boon. The loss of formatting (word wrap, capitalization) in DC.Description fields has been a problem that has been partially, but not fully addressed by recent improvements made by OCLC developers.

Editing. Because of our workflow, record creation or editing begins in DC view. The lack of a full-screen editor in DC view (until the release of October 27, 1999) has been a significant barrier for our project team, and we are very glad that OCLC has now developed a full screen editor (i.e., DC text area editing). We believe full-screen DC editing is mandatory in a CORC production system. Selectors and public services staff find the DC field-by-field editor we have used for most of the project, with its associated requirement for numerous "send and respond" interactions with the OCLC server, tedious and laborious. We expect to test the new DC editing functions, but we have not yet done so since we are close to the end of the research phase of our CORC project.

Diacritics and special characters. The Cornell library contains or provides access to materials in many languages. CORC's current input method for diacritics and special characters is thus inadequate, and it is very important that it be improved.

Duplicates. There are far too many duplicate records in the CORC database. We recommend OCLC act aggressively to resolve this problem and keep it solved.

Bad URLs. There are far too many bad URLs in the CORC database. Again, we recommend OCLC act aggressively to resolve this problem and keep it solved.

Searching. This is related to our concern about duplicates and bad URLs or similar URLs for the same site. We wonder if CORC searching could be adjusted and enhanced so that the user is more likely to discover already existing records, or if there are other steps that could be taken to help users find and/or avoid inputting duplicates.

Scorpion-generated subject headings. While the tool has tremendous potential, we believe that it needs more development. We particularly like the links from Dewey numbers to LC subject headings (since we don't classify in Dewey). If the Scorpion tools are installed in a production version of CORC, use of this tool in the creation of new records should continue to be optional. We would like it to be possible to apply the Scorpion tool to existing records as well. However, in the CORC database, the presence of so many inappropriate, sometimes laughably bad, Subject.DDC-Scorpion fields is a disservice. We recommend that they be systematically removed from at least the NetFirst records before CORC becomes a production system. Users of the tool should be strongly urged to remove inappropriate Scorpion-generated fields before submitting them to the database, leaving only appropriate ones on the completed database record.

Create multiple records in catalog. We have found this function very useful in our mini-project to catalog selected resources from ICE (Internet Connections for Engineering). We have several enhancement suggestions: (1) add an option to exclude local links, like the one available for the creation of a pathfinder; (2) make the system searching of links more sophisticated, so existing CORC records are more likely to be found; (3) instead or—of in addition to—listing the URL in the "Create Records" response list, list the title of the resource.

Constant data. It would be helpful if contributing libraries could create multiple constant data forms rather than just one. We would also like to be able to apply constant data to existing as well as new records. Catalogers would like to be able to use constant data to apply, for example, standard variable field notes describing restricted access or system requirements to existing records. It is also desirable to be able to create a constant data template for just the fixed fields (e.g., 007). We wonder, too, whether it might be helpful to develop a constant data feature for working in DC view. In our workflow, for example, the selector types a Description.Note field "Selected by [name]" into each DC record before placing it in the in-process file (once the record is completed, the note is removed). It would save repetitive typing if the selector could put this field into a constant data record which could be applied to either a new or existing DC record.

Export. The current functionality for transferring a MARC record from CORC to our local system is, to use the technical

term, "clunky" and needs to be improved.

Authority control. At Cornell, catalogers do authority wo

rk associated with the creation of new bibliographic records in our local system and in one of the utilities. We contribute our new NACO records through RLIN. Much of our remaining authority work is done post-cataloging, with the help of WLN. So, while we find the many authority control features built into CORC impressive, the workflow we have used so far in CORC requires little use of them. We have made use of the ability to "control" a heading in DC and MARC view and would like to see the current functionality expanded so the system can provide more help with the verification and correction of LC subject heading subdivisions.

Pathfinder functions. Further develop these, as noted earlier in this report.

[Back to "Recommendations" Sections](#)

[Back to Top](#)

Last modified 1999-12-11

ksc10@cornell.edu

[StaffWeb Index](#)