Testing the CORC System as a Tool for Managing Agricultural Information

Stuart R. Gagnon, M.S.L.S.
Science and Technology Services Team, University of Maryland Libraries, University of Maryland, College Park, MD 20742-7011, sgagnon@nal.usda.gov

Joseph R. Makuch, Ph.D.
Water Quality Information Center, National Agricultural Library, Agricultural Research Service, U.S. Department of Agriculture, 10301 Baltimore Ave., Beltsville, MD 20705-2531, jmakuch@nal.usda.gov

Sherry L. Davids, M.S.L.S.
Cataloging Branch, National Agricultural Library, Agricultural Research Service, U.S. Department of Agriculture, 10301 Baltimore Ave., Beltsville, MD 20705-2531, sdavids@nal.usda.gov

Written for presentation at the
2002 ASAE Annual International Meeting/ CIGR XVth World Congress
Sponsored by ASAE
Hyatt Regency Chicago
Chicago, Illinois, USA
July 28-July 31, 2002

Abstract. An important function of all libraries is to catalog information which can later be located and accessed by people who need it. Agricultural libraries are not exceptions. As the World Wide Web has emerged as an important and vast repository of agriculture-related information, so too has the need for a systematic mechanism to catalog (describe, identify and manage) electronic information for easily locating specific resources. To help address this need, libraries are expanding their information management techniques with a distributed system for metadata entry with a centralized database. Online Computer Library Center (OCLC), a leading global library system cooperative, has created the Cooperative Online Resource Catalog (CORC), a Web-based system model that facilitates metadata creation for Web resources selected and evaluated by libraries across the country. In 2001, the National Agricultural Library (NAL) began a pilot test of the CORC system to determine how it fits with the library’s current information management activities, such as producing AGRICOLA — a database of citations to the agricultural literature created by NAL and its cooperators. This paper describes NAL’s experience with the CORC system and how it may benefit producers and users of agricultural information.

Keywords. information systems; information; Internet; reports; journal articles; water quality; Water Quality Information Center; National Agricultural Library; Online Computer Library Center; Cooperative Online Resource Catalog
Introduction

The Cooperative Online Resource Catalog (CORC) is an online system containing databases for collecting, organizing and searching descriptive metadata to aid the processing and improve access to the electronic resources they represent. With properties inherent in electronic documents, such as the availability of multiple formats and versions, librarians were faced with difficulties in describing these new types of "holdings" (books, journals, and other resources owned by a library). Electronic holdings include those that are "born digital," items formerly available in paper format and now published only electronically (e.g., certain government documents and many serial publications, such as U.S. Department of Agriculture’s Oil Crops Outlook), and documents converted in digital preservation projects. CORC serves as one solution to the challenge of organizing electronic resources.

Hayes and Larson (2001) outline CORC benefits, including the opportunity for cooperation among libraries, cost effectiveness due to savings in intra-organizational software development, and CORC interface access to classification schemes useful as controlled variables in specific metadata fields. CORC provides the opportunity for active participation in selection by subject specialists; the dynamic creation of preliminary bibliographic records; the integrated development of subject bibliographies, called "pathfinders," both static and dynamically generated; managing and tracking Uniform Resource Locator (URL) changes; and the use of alternative metadata. CORC is a response to the needs of libraries collecting electronic resources.

Libraries and CORC

CORC is a distributed information system following the example of most online public access catalogs (OPACs) available from libraries. In this common model, libraries work cooperatively by entering bibliographic, or metadata, records to a larger cataloging system and thereby sharing workloads in a distributed manner. "Machine-readable bibliographic record[s] consist of fields," (Bibliographic Formats and Standards, 1996) as do metadata records. Therefore, for the purpose of this paper, metadata records representing electronic resources and the more traditional paper resources are termed "bibliographic records," even though the concept implies the description of more traditional written works.

Traditional books, the physical objects, held by several libraries are originally cataloged once, since bibliographic records associated with each book are accessible to all participating libraries throughout the larger system. Publishing organizations, government agencies among them, have made strides in providing more electronic versions of documents, articles, reports, papers and other publications as remotely accessible electronic files. Such resources can also be published directly to the Web and not ever become available through traditional channels. This situation presented a challenge to the library world and was a departure from the way information objects were described. CORC began as an initiative to help librarians deal with new aspects of describing remotely-accessible electronic resources on the Web (Ballard, 2001).

Traditional Library Systems

Librarians and the field of library science support the collection of and access to resources useful to their patrons, or users. Throughout its history, library science has developed systems, including electronic databases, to help provide better access. Libraries have traditionally developed organizational systems to provide assistance to patrons in their search for
information. These information retrieval benefits are seen as added value to the work of libraries and information centers. In the past, library information retrieval systems have been comprised of various types of physical catalogs, such as card catalogs, in specific arrangements referencing collection locations of represented books. The current extension of this model is the OPAC.

The accession process in libraries for traditional resources consists of selecting and then acquiring a physical object (a book, for example). The next step would be to describe the object prior to addition into the physical collection so that successful recall would be ensured.

Cataloging in libraries consists of describing a number of "handles" through which patrons can obtain items related to their specific search parameters. An extensive set of bibliographic record elements with corresponding syntax rules are necessary in order that the catalogs provide access at optimum function. [The bibliographic record fields used by libraries in the U.S. and Canada and known collectively as the MARC 21 (MACHINE-READABLE CATALOGING) set are maintained by the Library of Congress (2002). The corresponding rule set, called the Anglo-American Cataloging Rules (AACR), traces its roots to the 19th century and printed union catalogs.]

Library selection processes for paper-based material often involve purchasing the items recommended, or selected, by subject specialists such as bibliographers or subject liaison librarians. This can include the review of materials available through traditional book and journal publishing outlets. Electronic resources, especially the freely available ones, are often not made available through these sources.

As with free traditional material, freely accessible electronic resources are likely to lack adequate marketing efforts. CORC answers a need to assist subject specialist librarians, who are exposed to these resources, to identify and preliminarily describe these less accessible items. One of the world's largest shared-cataloging organizations, OCLC, developed the CORC system for its members with this need in mind.

**OCLC and WorldCat**

CORC is offered by the Online Computer Library Center, Inc. (OCLC) as a solution to cataloging electronic resources. OCLC began its distributed library catalog system, shared by Ohio libraries, in 1971. The system now includes 41,000 libraries in 82 countries (About OCLC. [2002]). OCLC's nonprofit status has kept this membership-led organization responsive to the needs of libraries in performing tasks for which there is a demand but no government or private service providers (Kilgour, 1966). OCLC sustains library cooperation through funding for services provided to member libraries. The organization works to increase the availability of library resources, to reduce information costs and to further scholarship and education in librarianship (Smith, 1998).

"A library, a group of libraries, or a national library network can be viewed as an information transfer system. A single library has at least four major subsystems: (1) an information store; (2) communications; (3) real property; and (4) users..." (Kilgour, 1966)

Frederick Kilgour, founder of OCLC, recognized the importance of periodic revision of library systems, bringing the design of the information system closer to the actual work of the library to connect the user to information. The earliest OCLC systems were developed to help libraries describe their holdings on representative cards in a card catalog system. This is where the term cataloging comes from, used for the process of creating card representations of the actual information entities (books, journals, etc.). In most of today's libraries, card catalogs have been
replaced by OPACs. Cooperative library cataloging has been the norm for over 30 years due to its greater efficiency and cost saving benefits. Libraries around the world participate in OCLC’s WorldCat system and contribute their work to this larger union catalog, a cooperative catalog with over 40,000 member libraries.

CORC is an effort to rework the principles of WorldCat in order to improve the control of online resources. CORC began in 1998 as an OCLC research and development effort. The project was preceded by the InterCat project, an earlier investigation of cataloging Internet resources (CORC Builds on Knowledge Gained from InterCat and NetFirst, 1998). CORC moved from research project to the status of OCLC production system in 2000, as stated in Technical Bulletin 239: Integration of the OCLC Cataloging Service and CORC (2000). The history and design of the CORC project is outlined in greater detail by Hsieh-Yee and Smith (2001a and 2001b) and Hickey (1998).

**NAL’s Interest in CORC**

The main reason for NAL’s participation in CORC is to improve access to agriculture-related electronic resources by identifying, selecting and cataloging these resources. Although NAL has been cataloging electronic resources since 1995, it was felt that the number and quality of resources added to AGRICOLA (NAL’s online catalog) could be increased by greater participation of NAL subject-matter specialists, such as information center staff, in selecting these resources. CORC seemed an ideal tool to facilitate this effort because the CORC input system, designed for use by individuals who are not experienced users of cataloging systems, can be learned quickly and without extensive training. Also, the ability to create pathfinders (electronic subject bibliographies) using CORC is useful to NAL’s subject matter specialists.

In addition, NAL recognizes the importance of maintaining links to cataloged resources. Link maintenance in CORC is facilitated by OCLC software that notifies libraries with holdings attached to a record that the record contains a broken/redirected URL.

**Impact of OCLC’s Global Strategy**

According to OCLC, access to regular cataloging functions will become fully Web-accessible in July 2002. By the end of 2003, Passport, the company’s terminal-emulation software, will be completely phased out. The new OCLC Cataloging interface will integrate CORC functionality and appearance (OCLC’s Global Strategy: Frequently Asked Questions, 2002 and OCLC Cataloging and Metadata Services: Guide to Migration, 2002).

So, a third reason for NAL’s interest in CORC is OCLC’s upcoming introduction of the new, browser-based interface for its cataloging and metadata services. With CORC’s browser-based interface, NAL believes that cataloging staff participating in CORC will have the advantage of advance training and experience prior to the upcoming phase out.

**CORC and the NAL Metadata Initiative**

Standard metadata field sets, such as MARC 21, have been developed to help catalogers create more uniform data for use in library information retrieval systems. CORC follows the MARC 21 model and provides options to save records in the Dublin Core (DC) format (Figure 1a), a leading metadata standard aimed at electronic discovery (Dublin Core Metadata Initiative, 2002), and Resource Description Format (RDF), a standard metadata framework for use with XML (Figure 1b). For a description of RDF and its uses, see Berners-Lee et al. (2001).
Figures 1a and 1b. Dublin Core and Resource Description Format record versions.
In anticipation of widespread proliferation of alternative metadata schemas, the NAL Metadata Task Force created a template for standard NAL metadata following the Dublin Core standard (NAL Metadata Task Force, 2001). The purpose of the template is to provide guidance for creating metadata to describe both electronic resources created within the Library and digital objects created by NAL from print objects. RDF functionality and XML-based systems are seen as the next steps.

**CORC Pathfinders**

NAL collects resources relating to all aspects of agriculture and related disciplines, with emphasis on certain subject areas including water quality, alternative farming, animal welfare, food and nutrition, technology transfer and rural information.

The CORC project would involve identifying and selecting electronic resources for CORC’s Resource Catalog (RC) database and creating information guides, called *pathfinders*, that consolidate links to electronic resources relating to water quality.

Pathfinders have been around for a long time. The pathfinders familiar to most of us are the lists created by library staff to help students, researchers and others conduct research on a specific topic. Usually prepared with a particular library’s holdings in mind, pathfinders provide a topical guide to available resources such as books, bibliographies, indexes and abstracting services, directories and encyclopedias.

CORC pathfinders follow the same general principles. They are online subject bibliographies, usually containing annotations, directing users to electronic resources by means of links that take them directly to the resources. However, CORC pathfinders are not limited to resources located within a specific library. CORC pathfinders are created online from bibliographic records for electronic resources found in the CORC database, to which libraries from all over the world have contributed. The quality of the electronic resources cataloged in CORC is high, because the resources have been selected and evaluated by librarians. And although the bibliographic records in the CORC database are for electronic resources, creators of pathfinders are free to include references to resources not in the CORC database and in any format.

An important feature of CORC pathfinders is the ability to create “dynamic” pathfinders. These pathfinders contain built-in search strategies that locate newly cataloged resources, thereby keeping the pathfinder up to date.

**NAL CORC Pilot Project**

NAL’s interest in CORC, expressed by the Cataloging Branch, began in early 1999. The branch wanted to launch a pilot project to explore how NAL might use the CORC system to improve its cataloging of electronic resources. NAL’s Water Quality Information Center (WQIC) was interested in participating in the project because the center and the Cataloging Branch had been working since November 1998 on a project to get freely available, online documents related to water and agriculture cataloged in AGRICOLA. A system was established whereby the center identified appropriate online publications and sent the titles, with access information (URLs), to Cataloging. There, the publications were reviewed again to ensure that they would fit within the NAL collection development policy, and then were given to a cataloger to create bibliographic records for AGRICOLA. Makuch and Hamilton (1999) provide a fuller description of this project.

WQIC experiences gained in the project, coupled with interest in expected capabilities of the CORC system (described above), fostered the center’s desire to work with the Cataloging Branch in testing CORC at NAL. In addition, WQIC staff had participated in the NAL Metadata Task Force and in other standard metadata projects (Gagnon and Makuch, 2001). With these
efforts in mind, the NAL Cataloging Branch requested that WQIC, as an information center with experienced staff, participate in NAL's CORC project, since an understanding of metadata issues was essential.

Administrative issues between NAL and OCLC and other priorities at NAL delayed the start of the CORC pilot project until June 2001. Before beginning, staff who were to be involved with the project participated in a one-day training course on building pathfinders.

The pilot project had several goals:

- evaluate workflow in Cataloging for processing online documents
- continue to add online documents to AGRICOLA, test procedures and identify impediments in the information center/cataloging workflow and propose solutions
- determine how difficult/easy it is to develop pathfinders; evaluate the usefulness and maintenance requirements of pathfinders
- determine how the CORC system fits in with current and planned NAL operations and policies
- document experiences with the CORC system and communicate them to other units of the library
- develop procedures that can be adapted by other NAL information centers and cataloging staff to implement CORC

**Project Description**

For the NAL project, it was decided that WQIC staff would send ten titles each week to Cataloging for addition to CORC. As of March 2002, 230 titles have been processed. The pool from which these titles were drawn was an internal WQIC database of freely available, online documents related to water and agriculture that the center maintains and is constantly adding to. WQIC staff members collect subject-related electronic resources throughout all work processes. Collected resources are currently placed in a local ProCite database. ProCite software provides a bibliographic management tool. The ProCite database is part of a local database prototype. Resources are selected for CORC from this source. Relevant Web sites, e.g., sites of water-related research units of the Agricultural Research Service, were also included as potential additions to CORC.

A NAL collection development staff member was asked to screen the selected resources, removing items not within the scope of NAL's collection. A NAL cataloger was assigned the task of processing the CORC records officially selected through collection development. The cataloger would follow USMARC formatting and rules to bring the record to an acceptable descriptive quality for inclusion in the NAL OPAC. WQIC would also produce several pathfinders, including a dynamic one. The workflow is described below.
Figures 2a and 2b. Screens for searching CORC’s Save File and Research Catalog.
**Process for Subject Specialists**

Once an electronic resource is selected, the WQIC staff member searches the CORC Save File (Figure 2a) and Resource Catalog (RC) databases (Figure 2b) to determine if a bibliographic record already exists for the selected resource.

The Save File contains NAL's incomplete, or "in progress," records prior to their inclusion in RC, CORC's main records database. The existence of Save File records shows that the resource has been submitted by NAL for cataloging. This search helps avoid redundant work on a single resource.

If a record cannot be found in the Save File, the RC is searched. Three outcomes are possible:

- a record may be completely new to the RC
- a record may be included and cataloged by one or more other libraries (not NAL)
- a record may have already been cataloged by NAL

For the last outcome the decision is to drop the resource, since it is already considered part of the NAL CORC collection. [Incidentally, an earlier step --- searching the NAL OPAC --- could have brought us to that finding. However, because our OPAC currently contains several entry points and collections that multiply the number of searches necessary, such a step, at this time, is not cost effective.]

For the first two outcomes, NAL records are created and saved to the Save File. Records are saved with the status of "new" whether or not a bibliographic record represents an existing non-NAL or brand new resource. New or not, WQIC staff marks each acceptable record by saving it to the Save File. In cooperative cataloging, an institution will accept as its own existing records cataloged by other libraries in the cooperative, hence the reason for saving an existing record to NAL's Save File, which contains all preliminary records.

If a bibliographic record does not exist for the resource, WQIC staff will create a preliminary Save File record. Creating a new preliminary record involves several steps. Preparation usually involves opening at least two windows to help workflow. With CORC and the resource open in browser windows,

- enter the resource's URL in the text block on CORC's Create in Catalog form (Figure 3)
- select Create --- a split window will load with the document (bottom frame) and the new record (top frame) (Figure 4)
- edit the record to an expected description level (optional)
- save the record.

This process is quite simple and lends itself directly to creating better metadata by placing the incomplete record in the hands of library professionals skilled in enhanced descriptive techniques. Beyond this process, which is internal to CORC, the NAL project used an email message formatted to track record entry. The extra steps associated with the email procedures were time consuming, but useful for project analysis.

Metadata standards are interchangeable in CORC. Subject specialists entering records for resources can use the more intuitive Dublin Core (DC) fields; catalogers can easily toggle between DC and the MARC 21 formats. This makes for a flexible interface for both metadata experts and non-experts.
Figure 3. CORC’s form for submitting URL for extraction.

Figure 4. Record screen during automatic harvesting process.
Process for NAL Technical Services

Following the subject specialist’s work, CORC contains a preliminary record for a selected resource, parts of which contain automatically generated information. Some records will contain a number of harvested metadata fields while others, only a few fields. This is usually dependent upon the amount of metadata already existing in the resource, or object, itself. For some objects, such as a certain type of portable document format (PDF) file, very little descriptive information can be extracted by CORC. For these resources, the subject specialist must type supplemental metadata into their corresponding records, with title information (MARC 21, field 245) and URL (MARC 21, field 856) as a minimum.

On the other extreme, Figure 5 shows an example of a CORC-created record generated from a metadata-rich resource. The original object was an HTML file which contained extensive embedded metadata in "meta-tags." The object’s metadata was automatically harvested by the CORC system and inserted into the record, requiring little or no entry by the subject specialist.

![Figure 5. Metadata-rich harvested record.](image)

On a weekly basis, NAL Collection Development Committee staff review preliminary records in the CORC Save File with a status of “new.” To aid in this process, records can be retrieved in the system by status level (Figure 6).

A NAL collection development librarian screens recommended selections, comparing them with the library’s collection development policy guidelines. All records for resources that fall within NAL collection development policy guidelines are considered appropriate to be cataloged into NAL’s holdings. In CORC, once a record for a recommended selection is accepted, collection development staff re-save the record to the Save File with a status of “in process.”

The NAL cataloger then refines bibliographic records, using MARC 21, for all resources with a status of “in process” that have not been cataloged by another institution. For records with a status of “in process” that have been cataloged by another institution, NAL’s code will be added.
to those records and possibly enhanced. Fully cataloged CORC records are then added to OCLC's WorldCat database automatically within 24 hours. Once electronic resources have been identified, selected, and cataloged, the bibliographic records are also exported to AGRICOLA, NAL's online catalog.

Because of our work with CORC and close evaluation of electronic resources, the selection process has been enhanced. One outcome was the revision of the Electronic Resources Selection Policy, Addendum No. 2 to the Collection Development Policy of the National Agricultural Library. The new version of the policy allows for more flexibility in the selection process. The policy now permits additional classes of Web resources.

**Pathfinder Process**

Pathfinders are information guides built with the CORC system (see description above). With regard to creating WQIC pathfinders, staff members gained experience creating both a static pathfinder and a dynamic one. Static pathfinders (Figure 7) include resources selected, or "tagged," from the RC; collected Web resources not cataloged to CORC; and resources in other formats, such as paper, computer disks and video tape. Dynamic pathfinders (Figure 8) rely upon queries pulling results from live searches of the RC. A pathfinder may have both characteristics. Figure 9 shows a finished pathfinder.
Figure 7. Editing a static pathfinder.

Figure 8. Dynamic pathfinder query-building.
The creation process for pathfinders follows the model of fielded entry for title, description, tagged records, (Web) link or query. Searching and saving processes follow the RC model in which all pathfinders are searchable and unfinished pathfinders remain in the Pathfinders Save File. Cloning is possible at any stage. Through another process, the CORC system assigns URLs so that pathfinders can become available as links from any Web site.

All CORC entities are editable, allowing changes to both dynamic and static pathfinders and metadata records to be done internally to the system, thus editing the output of each information product. Limitations were apparent in both dynamic and static pathfinders. Unwanted descriptive information related to the use of specific fields for describing physical objects was sometimes pulled into pathfinders. Control of formatting was not refined enough to improve the situation.

Familiarity with data input systems is helpful for the process of building pathfinders although not necessary because this process is straightforward. However, learning the system to create pathfinders with high-quality content and formatting requires practice. No WQIC pathfinders have been made public yet, so comments on their usefulness, popularity and related issues cannot be made. OCLC does not currently charge for accessing pathfinders, but may in the future.

Conclusions

NAL has the mission of ensuring and enhancing access to agricultural information for a better quality of life. One way to fulfill this mission is by maximizing access to information through collaborative efforts and utilization of technology (USDA, 1994). As online information has become more common, NAL, like all libraries, has been challenged to find ways to manage digital materials while continuing to improve management of hard copy items.
Participation in the CORC project has caused NAL to revisit its Electronic Resources Collection Development Policy; this policy has been revised to permit the inclusion of resources that consist entirely of collections of links because of their usefulness in creating pathfinders. The project, involving a technical services unit (Cataloging Branch) and a public services unit (WQIC) with distinct and synchronous day-to-day responsibilities, pointed out the need to establish feedback mechanisms between units to ensure efficient and effective workflow. The next step in the project will be to share our CORC experiences with NAL management and staff and help fine tune workflow and operations issues if the system becomes widely utilized at NAL.

According to OCLC’s Global Strategy, cooperating libraries around the world will be shifting to a Web interface based upon the CORC system. Constructed to promote more collaboration within libraries, CORC is a proving ground for inter-departmental, solution-based communication. Solving challenges in CORC may impact any library's transition to the new OCLC interface and through other changes.

As more subject specialist librarians in university settings become aware of CORC’s capabilities, agricultural colleges can assist in recommending appropriate electronic resources for CORC through library liaison relationships. In turn, this makes the resources more accessible to the schools’ researchers and students. A basic understanding of CORC on both sides can help drive the process.

As libraries face coming challenges, internal collaboration and external liaison activities may be enhanced by working with a system like CORC to better manage electronic and, by extension, traditional library resources.

All screen captures were taken from OCLC’s Cooperative Online Resource Catalog and used with permission from OCLC, Inc. CORC is a registered trademark of OCLC Online Computer Library Center, Inc.
References


Bibliographic Formats and Standards, 2nd Ed. 1996. 1.1 Machine-Readable Bibliographic

CORC Builds on Knowledge Gained from InterCat and NetFirst. 1998. OCLC Newsletter
Apr. 2002.

Dublin Core Metadata Initiative. 2002. Dublin Core Metadata Initiative (DCMI),

About RDF. World Wide Web Consortium (W3C), http://www.w3.org/RDF/FAQ.

Specialized Terminology Schema for a Web Database at the National Agricultural
Library" (paper presented at the 16th annual meeting of Computers in Libraries,


Hickey, T. 1998. Cooperative Online Resource Catalog Explores Uses for Catalog of Internet
Resources. OCLC Newsletter September/October (235): 15-19,

Part I. OCLC Systems & Services 17(3): 133-140.

Part II. OCLC Systems & Services 17(4): 166-177.

Kilgour, 17-20, ed. L.L Yoakam. Dublin, Ohio: OCLC,


Publications on the World Wide Web" (paper presented at the American Water
Resources Association, Symposium on Water Resources and the World Wide Web,


Computer Library Center, Inc.,

16


