Notes on Operations

Automated Access Level Cataloging for Internet Resources at Columbia University Libraries

By Kate Harcourt, Melanie Wacker, and Iris Wolley

The explosive growth of remote access electronic resources (e-resources) has added to the workload of libraries' cataloging departments. In response to this challenge, librarians developed various ways of providing access to electronic collections, but few dealt with the processing of free remote access e-resources, such as electronic books, Web sites, and databases. This paper will consider the various approaches taken by cataloging agencies to process Internet resources in all formats. It will then go on to describe Columbia University Libraries' approach to cataloging free Internet resources using a combination of selector input data, an automated form able to convert the information into MARC records, access level records, and cataloging expertise.

The cataloging of remote electronic resources (e-resources) has become a fact of life in the cataloging units of most libraries. Since the emergence of the Internet and remote e-resources in the 1970s, cataloging rules have had to be continuously adjusted to accommodate new developments. The increasing demand for access to online resources via library catalogs or library Web sites has also added to catalogers' workloads. This paper contains a literature review describing libraries' approaches to provide access to online collections, and introduces Columbia University Libraries' (CUL) solution for handling the cataloging of free Internet resources. The CUL approach combines selector input, an online request form with underlying programs converting data into Machine-Readable Cataloging (MARC 21) format, access level records, and a final review by cataloging staff.

The New Cataloging Environment

In the 1990s, with the growing popularity of the Web, more and more indi-

viduals and corporate bodies created their own Web sites and made their publications available online in addition to, or even instead of, their print counterparts. Publishers saw a marketing opportunity and quickly began to create and publish documents in electronic format. Commercial vendors promoted online over print counterparts either by using a pricing model that made continuing print subscriptions extremely expensive, or by discontinuing the print version entirely. Users and public services librarians then clamored to see remote e-resources in libraries' online catalogs, and technical services staff had to find ways to keep up with this new and growing workload.

This challenge is likely to increase even more in the future. On October 10, 2005, the BBC reported: "In its October survey, Netcraft [a monitoring firm] found 74.4 million Web addresses, a rise of more than 2.68 million from the September figure."[1] Also in October 2005, the "Six Key Challenges for Collection Development" presented at the Janus Conference outlined two goals that, if implemented, would impact e-resources cataloging immensely: the digitization of all

Kate Harcourt (harcourt@columbia.edu) is Head, Original Serial and Monograph Cataloging, **Melanie Wacker** (mw2064@columbia.edu) is Cataloger, Original Serial and Monograph Cataloging, and **Iris Wolley** (iw2117@columbia.edu) is Integrating Resources Cataloger, Original Serial and Monograph Cataloging, all at Butler Library, Columbia University Libraries, New York.

holdings of North American research libraries retrospectively as a national project, and the shift to purchasing electronic-only items when acquiring new publications.[2] As enormous amounts of information become available online, either free or through paid subscription, librarians have to tackle the ever growing task of how to select, provide access to, and manage all of these resources.

The number of cataloged non-serial remote access e-resources in Columbia Library Information Online (CLIO), the online catalog of CUL, jumped in just one year (2004 to 2005) by 359 percent, from 45,492 to 208,680. Although this number includes purchased records as well as those cataloged in-house, it nevertheless illustrates the growing demand for bibliographic access to information in electronic form. A substantial backlog of national and international online government publications existed, and the catalogers could not begin to analyze large sets of e-book collections or databases that contained other valuable resources. Selectors requested cataloging for free Internet resources using an online request form, but the requests often took a long time to fill. Paid e-resources were given priority and other e-material, by necessity, was relegated to a time-available basis. In 2005, an existing original cataloging position was redefined to include cataloging Internet resources. Even with this additional help, Columbia's original cataloging department could not keep up with the demand. Another approach had to be found.

The three staff members most deeply involved in e-resource cataloging formed a Work Group with the goal of establishing a workflow that would enable them to provide timely access to new publications and to process the backlog. Searching for ideas in the library literature and on Web sites of other cataloging departments, the Work Group found that many other libraries provided an online form to request cataloging of free Internet resources.[3] Generally, those forms send information via e-mail to the cataloging department. While this made it easier for selectors to submit their requests, it did not help the catalogers to keep up with them.

Literature Review

The problem was already apparent in 1999 when Gorman posed the question "Can we afford full cataloguing?"[4] Gorman acknowledged the fact that full cataloging, although preferable to other bibliographic control options, is very expensive and labor intensive. At the time, he introduced the idea of applying full cataloging to e-resources of "lasting value" and to use a less expensive option—Dublin Core (DC) for others.[5]

What solutions have been applied to this problem in the cataloging world?

The revised version of the Program for Cooperative Cataloging's (PCC) *Report of the Task Group to Survey PCC Libraries on Cataloging of Remote Access Electronic Resources,* published in January 2004, five years after Gorman's article, provides some answers.[6] Even though the report states that 95 percent of libraries responding to the PCC survey did catalog this type of resource, "[it] is clearly an activity that has grown greatly over a relatively short period of time, and cataloging agencies are continuing to adjust."[7] The task force found that very few of the responding libraries used metadata schemas other than MARC, such as DC, but were planning to begin using them.

A workflow that followed Gorman's recommendation was described by Huthwaite in her article "*AACR2* and Other Metadata Standards."[8] In order to provide access to their free, non-serial remote access e-resources, the librarians of the Queensland University of Technology Library and Griffith University Library use full cataloging according to the *Anglo American Cataloguing Rules,* 2nd edition (*AACR2*) for some resources determined to be important, and a DC-based schema for all others. Short records are created by reference librarians via an online form.[9] This information is then converted into brief MARC records. In this approach, personal and corporate names are only accessible by keyword searching. While filling out the form, the reference librarians flag certain resources for full cataloging following their local guidelines.

Different levels of cataloging using *AACR2* and MARC, however, appear to be the most popular option among the PCC survey respondents. Many make use of full, core, and minimal level records depending on the material and the needs of their institution. In addition, in 2004/2005 the Library of Congress (LC) tested and introduced an access level record for Internet resources.[10] Libraries now have four levels of cataloging from which to choose, but no consistent approach on when to apply a particular level is apparent. This is still largely determined by local priorities. York University Libraries, for example, use minimal level cataloging for component parts of large collections and for Internet resources that are free with a print subscription.[11] Catalogers have the option of treating the e-resources as an added copy to the print counterpart if one is available. Everything else is being cataloged as full standard. In other organizations the level of access is determined by subject specialists.

For e-journals having an equivalent print counterpart, a CONSER policy in section 31.2.3 of the *CONSER Cataloging Manual* explicitly allows the options of combining the description of both versions into a single record or creating a separate record for the electronic version(s).[12] CONSER propagated this guideline as an acceptable policy that can be used instead of cataloging an e-journal

separately per *AACR2* and the *Library of Congress Rule Interpretations* (*LCRI*). *LCRI* section 1.11A and LC's *Draft Interim Guidelines for Cataloging Electronic Resources* allow for applying a similar single record approach for monographs.[13] The OCLC document *Cataloging Electronic Resources: OCLC-MARC Coding Guidelines* describes this approach for any format.[14]

One of the questions in a 2003 survey undertaken by the Cataloging Electronic Resources/Electronic Resource Display in OPAC Task Force of the Illinois Library Computer Systems Organization User's Advisory Group (ILCSO) focused specifically on the choice of single versus multiple records. Chen reports: "Comments from those responding to the survey leaned toward the single record method, but the decision to use a single record or multiple (separate) records for various versions of print and electronic titles had clearly not yet been settled."[15] The 2004 *Report of the Task Group to Survey PCC Libraries on Cataloging of Remote Access Electronic Resources* also found a large number of libraries using the single record approach in their catalogs for at least a portion of their e-journals and monographic online resources.[16]

Most recently, the PCC Standing Committee on Automation Monograph Aggregator Task Group listed in its *Functional Requirements for Electronic Vendor Records (FREVR) Final Report* the different e-book cataloging approaches currently in use in library catalogs.[17] This task group described both single and multiple record options. Separate records are being created "either describing the original e-book in the bibliographic record and referring to the original edition or describing the original edition in the bibliographic record and referring to the reproduction."[18]

E-resources are also made available to patrons through Web lists.

Those listings can be found on many library Web sites. Most libraries provide separate lists of e-journals, e-books, and databases, some in alphabetical order, others by subject. The respondents in the ILCSO survey were "almost universally presenting some portion of their electronic holdings on Web lists instead of, or in addition to, their catalogs."[19] The same was found to be so in the PCC survey, which reported: "Over 92 [percent] of libraries (83 of 90) provide access to remote electronic resources in ways other than cataloging on the local system. Of those, 78 [percent] (65) provide access on library [Web] sites."[20] Most of those Web listings are not maintained by catalogers. In her article "Web lists or OPACs," Anderson remarked that "for years, libraries have provided multiple and redundant access to 'new' media in the form of catalog entries (prepared by technical services librarians) and separately maintained lists (prepared by public services librarians)."[21]

Automated Cataloging Projects

Faced with the fact that none of these options seemed to solve the problem of keeping current with the workload, enterprising librarians began to think of ways to automate at least part of the cataloging process. They also discovered ways to use one data source to create both Web lists and MARC records to avoid the duplication of work done by catalogers and public services staff. Most projects of this type focused on e-journal cataloging. Anderson describes the approach developed by the Virginia Commonwealth University (VCU) Libraries in 1999.[22] Using vendor-supplied data, VCU created an e-journal database for journals in aggregator databases that was searchable on the libraries' Web site and, at the same time, was used to automatically generate minimal-level MARC records for journals that were loaded into the catalog.

A year later, at the IUG (Innovative Users Group) 2000 Conference, Jiras of the Rochester Institute of Technology reported his library's approach to cataloging e-journals in unstable aggregator databases.[23] Rollins, reporting on the process, wrote, "In a nutshell, one creates records from vendor supplied data, imports them into the catalog, and when the information changes or is out of date, one does it again."[24]

The Hong Kong Baptist University Library developed an e-journal computer program (EJCOP) to provide access to their e-journals holdings.[25] This project also focused on e-journals residing in unstable aggregator databases. Vendor lists and pre-existing MARC records were combined to form a single full MARC record for each full-text journal. The program was also able to convert the MARC record into HTML in order to upload the information to the e-journal list on the library's Web site. EJCOP also was used to facilitate record maintenance on a monthly basis.

Banush, Kurth, and Pajerek described the Cornell University Library version of automated e-journal cataloging.[26] The Cornell model employs the separate record approach, not just for print and online journals, but also for different electronic versions from various aggregator databases. Very brief bibliographic records are generated using vendor-supplied title and holdings data. The computer program then adds standard MARC and locally defined fields. These records are not output to the bibliographic utilities and lack some information traditionally considered to be important, such as controlled subject access, classification, and linking fields. The authors noted, however, that their approach enabled the library to provide timely title level access to all journals hidden in aggregator databases, to use this data for maintaining their e-journal Web lists, and to perform regular maintenance.

As these examples of automated cataloging projects show, the problem of keeping pace with the cataloging of e-journals, particularly those residing in large aggregator databases, has been addressed in a variety of ways. Much less effort has focused on how to automate the processing of non-serial e-resources, such as e-books, databases, and Web sites.

In 2001, the University of Florida established a nearly fully automated workflow for cataloging e-publications residing in the Extension Digital Information Sources (EDIS) database of the Institute of Food and Agricultural Sciences (IFAS).[27] A computer program, E-pub to MARC (E2M), was able to capture the necessary information from the electronic document itself through use of a Web crawler. A MARC converter then transcribed the data into a MARC record. Cataloging rules were followed and authority control performed. The records included summaries and contents notes, but lacked subject headings, classification, and added author entries. The MARC records were loaded into the local online catalog and into OCLC's WorldCat. The software was written for specific publications and depended on standardized HTML coding. The automatic processing of the IFAS publications using E2M ceased when the structure of the documents changed.[28]

The Library of Congress Bibliographic Enrichment Advisory Team (BEAT) recently introduced the Web Cataloging Assistant.[29] The cataloger copies a specific publication's uniform resource locator (URL) into the program, which retrieves bibliographic information directly from the resource and adds generic information. The software creates a MARC record from this data and sends it to LC's Voyager cataloging client. Catalogers update the records manually and add subject access and other necessary information. The Web Cataloging Assistant needs, just as E2M did, a "predictable and consistent layout of the bibliographic data."[30] It is, therefore, primarily used for works in specific monographic series that provide such a reliable structure.

In the *FREVR Final Report,* the PCC Standing Committee on Automation Monograph Aggregator Task Group recommended machine-generated catalog records by vendors as a way to provide title-level access to e-books residing in large aggregator databases.[31] While this would solve much of the problem, many other publications that are not the responsibility of any vendor or publisher are available online. These include international government and nongovernmental organizations' reports or Web sites. Libraries need to find ways to provide access to all this information.

E-Resources at CUL

CUL's struggle to catalog and provide access to electronic materials mirrors experiences in libraries worldwide. In February 1995, the Cataloging Department hired an e-resources/metadata cataloger to provide full cataloging, including serial holdings, for e-resources in all formats. Catalogers and managers discovered that creating and maintaining accurate e-journal holdings data was impossible and that, even with the addition of a bibliographic assistant, the Cataloging Department was not staffed to handle the volume of new digitized titles in an expanding array of formats.

In the same year, CUL sent a cataloger to OCLC to study the feasibility of using DC for certain categories of material. After much discussion and participation in the early stages of the Cooperative Online Resource Catalog (CORC) project, managers decided little would be gained through incorporating DC into Columbia's existing cataloging activities.

CUL next began to explore ways to obtain vendor-supplied cataloging but was discouraged by the quality and scarcity of records. In 2002, Columbia cataloging administrators and the CONSER Coordinator at LC began working with Serials Solutions to develop specifications for creating CONSER-based e-journal cataloging for journals in aggregator packages. Serials Solutions searches the CONSER database for a matching bibliographic record. When a record for the e-journal does not exist, Serials Solutions creates an e-journal record by extracting agreed-upon elements (if available) from CONSER print or microform records. When no CONSER record exists, Serials Solutions creates records based on data from Thomson Gale, Ulrich's Periodicals Directory, Serials Solutions' own in-house catalogs, and other sources. In this way, Serials Solutions provides customers with 100 percent coverage of titles and holdings for serial aggregations. This success encouraged CUL selectors to seek additional sources for vendor-supplied MARC records in all formats. By 2006, CUL had obtained as many MARC records as possible for paid e-journals and non-serial e-resources, including U.S. government documents.

In addition to cataloging paid resources and titles within aggregations, CUL made an attempt to catalog free Internet resources. Selectors sent notifications using an e-mail form informing the cataloging staff that a resource should be cataloged. Many of the requests came from selectors in the Area Studies Department collecting materials from Latin America, the former Soviet Union, and Southeast Asia as well as from selectors in the sciences.

An even larger volume of requests came from CUL's government information librarian. A U.S. federal documents depository since 1882, the Libraries have subscribed to the MARCIVE service for government documents since August 1994. MARCIVE, however, does not provide MARC records

for Web sites; thus, the Cataloging Department received requests to catalog these and address other gaps in vendor coverage, including publications from foreign governments and nongovernmental organizations. The Cataloging Department gave these latter requests lower priority than paid resources because of volume and staffing constraints. All e-resources, paid or free, were cataloged at full, PCC, or CONSER levels.

Another pressure for the Cataloging Department arose when CUL began several projects to extract metadata from MARC records for remote access e-resources in order to create specialized interfaces and e-resource lists outside of OPAC, usually by form (e.g., e-journals) or genre (e.g., reference tools and indexes). These lists are located at CUL's E-Resources Web site at www.columbia.edu/cu/lweb/eresources. The cataloging records used in these projects require special fields and procedures, necessitating extra time and expertise on the part of the cataloger. Metadata are harvested from bibliographic records and loaded into the enterprise SQL system (IBM's dB2) that acts as a "master metadata file," enabling real time searching and subject browse functionality. Subject access is achieved through LC call numbers extracted from the 050 field and mapped into Columbia's Hierarchical Interface to LC Classification (HILCC).[32]

After most of the libraries' e-resources were cataloged using vendor-supplied records, and a routine workflow was developed to handle the bibliographic records used for the extraction of metadata, staff members could consider how to provide bibliographic access to those not being addressed. In addition to the free e-resource categories previously identified, access was not being provided to component parts of paid databases. Selectors in many areas demanded better access to resources buried within large databases and Web sites. In

addition, when paper subscriptions to many monographic series had been canceled in 2004, staff members were not available to catalog the electronic versions selected to replace them.

Access Level Records

The Work Group investigated the possibility of adopting the access level record for remote access e-resources used at LC. In 2003, LC released an initial report recommending how bibliographic control and access for these types of resources could be accomplished.[33] One recommendation was a new type of record for a subset of Internet resources, one which would be rich in fields reflecting content and access and less full in descriptive fields. The record level developed by LC is an access level record that uses *AACR2* and LC Subject Headings. The content designation conforms to MARC 21.

Delsey's report *Defining an Access Level MARC/AACR Catalog Record* described scope, methodology, and guidelines that help define this record level.[34] Appendix A in the report provided a core data set containing user tasks and evaluations made regarding importance of use of various fields and subfields. In early 2005, Reser reported on test results of access level use.[35] Of special interest in this report are the results of cataloger time spent creating full records versus access records and the number of authority records not created.

Access Level Records at Columbia

In mid-2005, the Work Group examined LC's access level model for cataloging Internet resources. Ensuing discussions centered on the core data set and LC's decisions for access level records contained in the revised Appendixes B and C of Delsey's

report.[36] The Work Group evaluated the usefulness of fields and subfields, and discussed subject analysis, main and added entries, and classification. Each member brought years of Internet resource cataloging experience to the discussion and determined that some descriptive fields were not necessary for resource discovery, did not add to description, and sometimes provided redundant information. Among the fields not used in CUL's access record are the 260 field, all 3xx fields, and most 5xx fields. Use of the 246 field is limited to variant titles readily available. Work Group members determined that cataloger judgment should be the most important guideline when using CUL's access record. The record contains a basic set of fields to which other fields can be added if catalogers judge them to be of value for resource discovery. LC guidelines were crucial in supporting the group's goal of providing access and streamlining the use of descriptive fields. Work Group members adopted many of them. Appendix A at the end of this paper provides a comparison of descriptive fields used by CUL and LC in access records.

Subjects, main, and added entries, and classification follow LC's guidelines found in Appendix C of Delsey's report.[37] Work Group members believed that these fields enrich access to Internet resources. Full subject analysis is applied to each resource using as many subject added entries and index terms as necessary. These include 600, 610, 611, 630, 650, 651, and 653 fields. Catalogers create SACO headings if necessary. Main and added entries are used when appropriate and include 100, 110, 111, 130, 700, 710, 711, 730, and 773 fields. CUL's access level guidelines support the creation of NACO records for those headings not under control. CUL selectors use the LC classification number contained in the bibliographic record for collection development purposes. CUL

catalogers therefore continue to provide subfield $a of the 050:_4: field in access level records for Internet resources. Subfield $b is used only when needed to complete the class number.

The Work Group decided not to test cataloging time between full and access level records. This was based on the assumption that the results from LC's testing would be similar at CUL. For the same reason CUL catalogers did not time access level record cataloging for comparison with those recorded by LC.

Catalogers began to use the access level record in July 2005. Selectors continued to use the same e-mail form as before and send printouts of Web resources to inform catalogers which titles needed to be included in the online catalog. During the next few months, catalogers noticed that they spent much less time finding information regarding publication data, first iterations, what terms should be used in the 246 $i, and other elusive descriptive information. They could concentrate on subject analysis and authority control. The backlog of printouts and e-mail forms was coming under control. The application of fewer fixed and variable data fields resulted in a more standard record for Internet resources.

Automated Cataloging of E-Resources at CUL

The Work Group had been interested in generating MARC records from a predefined source of information since the initial evaluation of access level records. Could a MARC record be generated automatically from some source of information about each Web resource? Toward the end of summer 2005, the group began discussing this possibility. One very important realization emerged from the discussions: the workflow involved in receiving automatically generated

MARC records would need to begin outside the Cataloging Department. Identification of Web resources for inclusion in the online catalog began with the selection process. Thus, the group decided that selectors would fill out an online form with data about the resource from which a MARC record would be generated.

The process of extracting data from the form needed to involve CUL library systems staff, as well. Library systems staff could not begin their work without a clear design for the online request form. The first step, then, was to define default codes and field content for the MARC record, which would be generated from the new online request form.

Designing the Automated Cataloging Form

The Work Group designed a new Internet Resource Cataloging Request (IRCR) form, in consultation with the Library Systems department. Library Systems staff estimates that consultations, design, and programming took thirty-five hours of staff time. The Work Group decided to make the form as simple as possible for selectors and public service librarians while at the same time obtaining sufficient cataloging data. Terminology for the different field labels was chosen in consultation with selectors in order to avoid cataloging jargon. The IRCR form (figure 1) is located on Columbia's secure server and selectors must authenticate by inputting their e-mail ID and password in order to access the form. The only required fields are title and URL. The selector has the option of including Alternate Titles (246), Authors (7XX), Description (520), Subject keywords (653), Part of Resource (773), and a free-text "Note to Cataloger." Selectors do not need to "sign" their requests. Instead, a field is automatically populated with the selector's unique University Network ID (UNI).

This field is captured during user authentication and allows the Work Group to contact the selector if there are any questions. It is also used for statistical purposes. Some selectors use their UNI as a keyword search to see what has been cataloged. After the selector submits the form, a review screen is presented (figure 2).

The selector can edit or click OK to submit. If "edit" is chosen, the selector using the online form is returned to the form populated with the data already entered so that it can be revised. The last screen seen by the selector after clicking OK is a confirmation notice that includes date, title of the resource, and an assurance that a bibliographic record will appear in the OPAC in three working days.

Practical Extraction and Reporting Language (PERL) and MARC-related PERL modules are used to generate the MARC records. A Common Gateway Interface (CGI) program written in PERL generates the form and processes the data submitted. CGI allows HTML pages to interact with programming applications. The program was developed by Gary Bertchume, Senior Library Systems Analyst at Columbia University, and is freely available upon request to the authors. Programming provides an automated, single platform, Web-based solution that allows for unpredictable selector input but guarantees output for the cataloging staff whenever a form is submitted. Completely automating this process required the use of centrally maintained Unix Web servers, programs, and scripts that could run unattended in that environment. Data input into the form are gathered in an accumulation file on the Web server each time a form is submitted. A shell script is run daily to:

- copy the day's input to a work file and reinitialize the accumulation file for the next day's input;

- process the work file using a locally developed program, which generates a file of MARC records using the variable data found in the work file combined with a set of specified default values. Editing is done to remove control characters (e.g., tabs or carriage returns), to trim extra spaces, and to make sure that the URL is well-formed; and
- post the file of MARC records to the secure Web server and send e-mail to cataloging staff to alert them that a new file is ready and to supply the pickup URL, which allows the cataloger to access the file. The e-mail to the catalogers includes a link to a text version of the file for preview and quality control.

Discrete files for each day's accumulation are exported to the Voyager Workfile or Import file depending on cataloger preference. The file name begins with "ircr," the file creation date, and a .bin extension. A file created December 1, 2005, thus would be named "ircr_200512010200.bin." Catalogers import the records one by one from the file

Figure 1. Internet resource cataloging request form

Figure 2. Review screen

into the Voyager cataloging client and edit them for final production. The automatically generated MARC records contain some fields that are

machine-generated through the IRCR form, and others that are supplied by the program. The coding for the fixed fields (Leader, 008, 006, and 007)

is entirely predefined and program supplied. Fixed fields are not edited by the cataloger, with the exception of language and, for PDFs only, the publication date (figure 3). This represents a further reduction of required fixed field elements from those used in CUL's access record. The Work Group decided to take this step to take full advantage of the automated record creation. Figure 3 shows the fixed fields as supplied by the program.

The variable fields corresponding to the IRCR form are only generated if the selector supplies data. Other variable fields are program supplied and contained in every record. Figure 4 shows an example of a MARC record before review by the cataloger.

To keep the form as simple as possible for the selector, certain compromises were made and the resulting record requires careful review in several areas. All submissions generate records in integrating resources format. Until June 2006, the records defaulted to monograph format. After the implementation of the new inte-

grating resources Leader and 008 field at OCLC, CUL's library systems staff quickly revised the form, proving that the new workflow would survive major changes in cataloging practice. Asking the selectors to differentiate formats did not seem realistic. If the cataloger determines the title is not an integrating resource, he or she must change the bibliographic level. Catalogers currently catalog serials to full standard. The Work Group plans to apply the access level model to serials later in 2007 when PCC and LC complete their charge to extend the model to serials.[38] The selector may or may not include initial articles, so the cataloger may need to adjust the 245 field for proper filing. The general material designation "electronic resource" is automatically supplied at the end of the 245 field and sometimes needs to be moved to the correct position by the cataloger if the resource title has a subtitle. The default for author is a corporate author with name in direct order (710 2), so the cataloger must retag personal names or adjust their

indicators. The summary (520) is often copied and pasted from the online resource so Unicode conversion problems sometimes occur. Figure 5 shows a completed catalog record.

Testing and Implementation

The IRCR form and cataloging workflow were tested by Work Group members before the form was made available to selectors, in two phases of testing between late September and late November 2005. The first test, done within the Cataloging Department, was to successfully generate MARC records from the information input into IRCR forms. Work files were created overnight and Work Group members were automatically sent e-mail messages containing two URLs—one for the records that would be saved to the Voyager import file and one for the text documents containing data from the IRCR forms. This test confirmed that MARC records could be generated from the IRCR forms, so the second test was implemented.

The goals for the second test were successful generation of large daily amounts of MARC records over a long period of time and successful cataloging workflow management. Participants included the three catalogers from the Work Group and a selector, who had taken part in the initial planning of the project and who was a regular contributor of e-resources titles under the previous request procedure. During October and November 2005, the selector submitted 147 records through use of the IRCR form. Each Work Group member was responsible for cataloging Internet resource titles for one week at a time on a rotating basis. At the end of this test phase, the Work Group confirmed that large numbers of

Leader			008	Fixed-length data elements
05	Record status = n		06	Pub. Status = c
06	Type of record = a		07-10	Date 1 = uuuu
07	Bibliographic level = i		11-14	Date 2 = 9999
08	Type of control = blank		15-17	Place of publication = xx
17	Encoding level = 3		18	Frequency = u
18	Cataloging form = a		19	Regularity = u
			21	Type of CR = w
			22	Form of orig. item = s
			23	Form of item = s
			33	Orig. alphabet = blank
			34	Entry convention = 2
			35-37	Language = eng
			38	Modified record = blank
			39	Cataloging source = d
006	**Fixed length data elements**		**007**	**Physical description fixed field**
00	Form of material = m		00	Category of material = c
05	Target audience = blank		01	Specific material designation = r
09	Type of file = d			
11	Government publication = blank			
Note: all other available positions are coded with fill characters				

Figure 3. Fixed fields supplied by the program

	Tag	I1	I2	Subfield Data
	040			‡a NNC ‡c NNC
	050		4	‡a ?
	245	0	0	‡a International Rescue Committee ‡h [electronic resource].
	246	1	3	‡a IRC International Rescue Committee
	500			‡a Viewed on ().
	520			‡a Founded in 1933, the IRC is a global leader in emergency relief, rehabilitation, protection of human rights, post-conflict development, resettlement services and advocacy for those uprooted or affected by conflict and oppression.
	653			‡a Refugees ‡a Humanitarian efforts
	710	2		‡a International Rescue Committee.
	852			‡x iw2117
	856	4	0	‡u http://www.theirc.org/
	965			‡a ircr

Figure 4. MARC record before review

records could be supported, and that management of the new cataloging workflow, including a one- to three-day turnaround time, was sustainable.

The Work Group's next major decision was whether the extra step of searching OCLC and potentially doing some cataloging there was necessary. Columbia University Libraries use OCLC as its primary source of cataloging copy and is an OCLC National Level Enhance library. The CUL corporate culture supports creating original records in OCLC and enhancing cataloging copy when necessary. Catalogers work in either OCLC or in the local system, depending on expediency and judgment. The Work Group was aware that LC opted not to search the utilities for copy before creating their access level records and wondered whether working only in the local system would be more efficient. The Work Group decided that catalogers would continue to choose where to catalog using the same criteria used for other CUL cataloging work. Influencing this decision were surpris-

ing amounts of cataloging copy found and commitment to NACO authority work, necessitating use of OCLC for name authority record creation and review.

After evaluating the second test's results, the Work Group decided to share the new process for submitting and cataloging Internet resources with selectors and other CUL librarians. Work Group members and the selector who was a participant in the second test presented a program on the new access level record and IRCR form at a selectors' meeting in December 2005. The presentation covered the IRCR form and its development, selector and cataloging workflow, and basic fields of the access level record. The overall response from the selectors was positive and, within days, selectors began to use the IRCR form.

Performance and Evaluation

The use of the IRCR form in combination with automated cataloging has

provided an answer to many of the challenges created by the explosive growth of electronic information. The CUL catalogers now have a tool to provide timely access to free Internet resources submitted by selectors for cataloging. The prescribed turn-around time is three working days, but, in most cases, the records are upgraded the next day. This has had an immense impact on the workflow of the three staff members involved in cataloging free Internet resources. Instead of trying to make time whenever possible, the processing of free Internet resources has become part of the daily routine. Previously, only paid subscription databases and electronic collections received this kind of attention. By sharing the cataloging process with the selectors and employing an automated cataloging technique, the catalogers are able to concentrate their time on the creation of subject headings, access points, and authority work.

Occasionally, selectors submit more than twenty requests a day. This reduces the time available to the affected catalogers for other tasks. Cataloging staff do not feel that other assignments have suffered, since they rotate weeks for cataloging the files of requests and help each other out when a "bottleneck" develops. If the daily workload continues to increase, the Work Group may rethink some of the workflow decisions.

The Work Group timed the original cataloging of non-serial e-resources using the IRCR form for several weeks. The average cataloging time, including authority work, was sixteen minutes per record. Another expe-

Figure 5. Completed catalog record

rienced cataloger processed a small sample of integrating resources and e-books as full standard MARC records without help of the electronic form. The resulting average cataloging time of 31.5 minutes substantiated the group's belief that great time savings had been accomplished. CUL catalogers feel that these time savings of 44 percent can be attributed to the combination of four factors:

- Access level records eliminate the need of searching for hidden information, such as date and place of publication, and corporate bodies.
- The automated form saves catalogers time spent on typing.
- Selectors providing summaries and keywords simplify subject analysis.
- Reliance on cataloger's judgment rather than on strict rules eliminates the need to agonize over decisions and provides catalogers with the freedom to add

additional information when necessary.

LC catalogers involved in the LC pilot project voiced mostly favorable opinions on the creation of access records, such as "a breath of fresh air," "provided summaries were a big benefit," or "elimination of redundancies."[39] CUL catalogers agree with all of them, and add that the automated form amplifies the advantages of access records. CUL's emphasis on cataloger judgment resolves possible limitations of those records. Between October 2005 and April 2006, 836 submissions were cataloged using the new method. The Work Group decided to include component parts of licensed e-resources into the workflow as well, reasoning that since the main resource already went through the acquisition process its component parts could be considered "free" and submitted along with other free remote access e-resources. This decision presented CUL catalogers with a tool to pro-

vide access to valuable resources previously hidden within large aggregator databases.

In July 2006, the Original Cataloging Department was able to report a 24.6 percent jump in cataloging production for the 2005/06 fiscal year. Cataloging managers attributed most of this increase to use of the IRCR form in combination with access level records.

One of the most rewarding outcomes of the project has been collaborative problem solving. Selectors often provide summaries, keywords, and added entries that they consider to be important. They also provide references to related print resources or suggest subject headings via the note field. Good communication between catalogers and selectors has become critical. The introduction of the IRCR form not only brought free Internet resources to the fore in cataloging, but also generated discussions in public services. The improved information exchange made it obvious to the catalogers and selectors that various problems arose repeatedly during the cataloging process, but were settled on a case-by-case basis. The CUL government information librarian, in consultation with other selectors and the Work Group, drafted a long-needed policy defining selection criteria for free Internet resources.[40] For instance, free and paid content are occasionally offered on the same site. The staff members involved in drafting the policy decided that these resources are cataloged only if they make that distinction obvious to the patron. Many resources require the user to register, usually by providing an e-mail address. The Work Group was concerned that

some sites might pose privacy and security problems, depending on the information requested. The policy now states that CUL should continue to provide access to this type of material if considered to be of great value. The Work Group agreed to include the registration requirements in a note (506 MARC field) in the catalog record to alleviate the privacy and security concerns. This "Restrictions on Access" note displays prominently in the OPAC.

The new workflow for free Internet resources has been a great success from the technical service point of view, but does it work for the selectors? In order to answer this question, the Work Group formulated a short survey and, in early March 2006, sent it to thirty-six selectors (see appendix B). Hard copies of the survey were also distributed in the selector area of the acquisitions department. Before the deadline of two weeks, nine selectors responded. Since, by this time, not all the selectors had chosen to select free Internet resources for addition to the catalog, the Work Group decided that the nine responses were sufficient to evaluate the use of the IRCR form.

The feedback was positive. Only one respondent preferred sending e-mail messages directly to the Cataloging Department. Four selectors had used the form, while five had not but were planning on doing so. The impact on their work was generally judged as positive. One respondent wrote, "I love it. It is such an efficient way to get the record into our OPAC. Without this, I would need to baby-sit each title through the process . . ." Another selector remarked that the new form and automated cataloging process "reduce paperwork, make tracking easier, and result in faster cataloging." The only criticism was a first impression that filling out the form might be a little more work for the selector compared to the previous submission process.

Five selectors judged the ability to track their submissions by using keyword searches and their UNI as important or very important. This feature enables them to make sure the resource was cataloged and gather their own statistics. One person stated that locating submissions by UNI is useful when handling reference questions; another used it to revisit certain sites to keep track of changes. The other four respondents either did not use this option or thought it to be useful but did not consider it to be essential.

The Work Group asked if the selectors considered the ability to contribute keywords, summaries, and other cataloging data as important. The replies ranged from "somewhat important" to "critically important." The respondents loved being able to make use of their specialized knowledge in their subject area to point out additional titles under which a particular resource might be known, or to bring out special aspects that might not warrant a subject heading but are useful for information retrieval.

One of the replies referred to the closer working relationship between catalogers and selectors:

> If I've already spent some time reviewing the site to determine whether it is worth adding to CLIO, then I have some knowledge of its content and that should be passed on to the catalogers so they don't have to start from scratch. Even if they have good reason not to use my suggestions, it seems useful to suggest them. It also helps if the sites are in languages that the catalogers don't work with. Finally, a summary may be helpful when the title of a site isn't very informative, and increases the likelihood of discovery through CLIO keyword searches.

Only one respondent felt that this was not crucial and thought that "cata-logers could handle the whole thing more efficiently and more consistently." This selector also remarked that, in his opinion, optional selector input of keywords and summaries had not been made clear.

The Work Group asked if access level records were considered to be sufficient, both from the selector and public services points of view, or if any important information was missing. Seven respondents were completely satisfied. The other two selectors found the new model to be adequate, but also remarked that "full is better." No respondent noted any specific data element thought to be lacking in the records.

The catalogers involved consider the feedback from the selectors as very crucial to their work. The selectors were all pleased with the one- to three-day turnaround time and found the IRCR form easy to use. Some had trouble locating it on CUL's networked e-resources Web site. The Work Group will address this last point in the future.

Based on the responses, the new workflow appears to be as much of an improvement for the selectors as it is for the catalogers. The government document librarian, who helped the Work Group during the implementation phase of the form, commented, "As of today (Mar. 6, 2006), I have had 428 items cataloged via the Internet Resources Cataloging Request Form. In my opinion, that represents a significant addition to the electronic research material now available to Columbia University students and faculty."

Conclusion

Online resources play a major role in today's information environment. Providing access to all types of e-resource collections is crucial. CUL developed an automated cataloging workflow for free e-resources—one that includes selector input into the

cataloging process, provides online cataloging forms, and automatically generates MARC records.

In the months since the successful implementation of the IRCR form, many ideas have surfaced on how this automated cataloging workflow could be extended to other library technical services areas. The Work Group also realizes that other libraries could adapt the form and the underlying program to their own needs and projects. The form could be customized to accommodate other types of materials, such as microfilms, analytics, or to provide bibliographic access to pamphlets in vertical files. It could be adapted to handle large projects without putting strain on existing professional cataloging staff. Cataloging data also could be put into a spreadsheet instead of the form. MARC records are generated in the same way. Whether using the form or a spreadsheet, the underlying programs can be easily customized to generate resource or project specific data such as a series, added entries, or notes.

Incorporation of techniques developed by the Work Group into other technical services departments and activities is a high priority for CUL. Librarians and managers are equally excited about opportunities to create quality records more easily. This new approach gives the cataloger more time to focus on subject analysis and authority control and gives patrons access to underserved areas of the collections.

References and Notes

1. "Web Enjoys Year of Biggest Growth," *BBC News,* Oct. 10, 2005, http://news.bbc.co.uk/1/hi/technology/4325918.stm (accessed Mar. 30, 2006).

2. Janus Conference on Research Library Collections: Managing the Shifting Ground between Writers and Readers, "Six Key Challenges for Collection Development (Original Document)" (Oct. 2005). http://janusconference.library.cornell.edu/?p=49 (accessed Aug. 18, 2006).

3. Examples include: University of Virginia Library Cataloging Services, *Library Cataloging Request Form for Adding Library Free Nonserial Internet Resources to VIRGO Online Catalog,* www.lib.virginia.edu/cataloging/policies/forms/web-req.html (accessed Aug. 24, 2006); Houston Cole Library, *Internet Site Cataloging Request Form.* www.jsu.edu/depart/library/graphic/catint.htm (accessed Aug. 24, 2006); UCSD Libraries University of California, San Diego, *Request to Catalog an Internet Resource.* http://tpot.ucsd.edu/Cataloging/coldev.html (accessed Aug. 24, 2006).

4. Michael Gorman, "Metadata or Cataloguing? A False Choice," *Journal of Internet Cataloging* 2, no. 1 (1999): 12.

5. Ibid., 13.

6. Task Group to Survey PCC Libraries on Cataloging of Remote Access Electronic Resources, *Report of the Task Group to Survey PCC Libraries on Cataloging of Remote Access Electronic Resources,* rev. (2004). www.loc.gov/catdir/pcc/archive/tgsrvyeres_final.pdf (accessed Jan. 16, 2006).

7. Ibid., 6.

8. Ann Huthwaite, "*AACR2* and Other Metadata Standards: The Way Forward," *Cataloging & Classification Quarterly* 36, no. 3/4 (2003): 87–100.

9. *Anglo-American Cataloging Rules,* 2nd ed., 2002 rev. (Ottawa: Canadian Library Assn.; London: Library Assn. Publishing; Chicago: ALA, 2002).

10. David Reser, *It's All About Access!* (June 2005). www.loc.gov/catdir/access/ala_erig.ppt (accessed Jan. 16, 2006).

11. York University Libraries, *York University Libraries Procedures for Cataloguing Electronic/Internet Resources* (Toronto: York University Libraries, 2001–). www.info.library.yorku.ca/techserv/eres.htm (accessed Jan. 16, 2006).

12. Library of Congress Serial Record Division, *CONSER Cataloging Manual,* 2002 ed. (Washington, D.C.: Library of Congress Cataloging Distribution Service, 2002–).

13. Library of Congress, Cataloging Policy and Support Office, *Library of Congress Rule Interpretations,* 2nd ed. (Washington, D.C.: Library of Congress, 1989–); Library of Congress, Cataloging Policy and Support Office, *Draft Interim Guidelines for Cataloging Electronic Resources* (Washington, D.C.: Cataloging Policy and Support Office, 1997). www.loc.gov/catdir/cpso/elec_res.html (accessed Mar. 30, 2006).

14. Jay Weitz, *Cataloging Electronic Resources: OCLC-MARC Coding Guidelines,* rev. (2004). www.oclc.org/support/documentation/worldcat/cataloging/electronicresources (accessed Jan. 27, 2006).

15. Xiaotian Chen et al., "E-Resource Cataloging Practices: A Survey of Academic Libraries and Consortia," *Serials Librarian* 47, no. 1/2 (2004): 163.

16. Task Group to Survey PCC Libraries on Cataloging of Remote Access Electronic Resources, *Report of the Task Group,* 9.

17. PCC Standing Committee on Automation Monograph Aggregator Task Group, *Functional Requirements for Electronic Vendor Records (FREVR) Final Report* (2006). http://platinum.ohiolink.edu/dms/frevrfinalreportbecky.pdf (accessed Feb. 7, 2006).

18. Ibid., 2.

19. Chen et al., "E-Resource Cataloging Practices," 175.

20. Task Group to Survey PCC Libraries on Cataloging of Remote Access Electronic Resources, *Report of the Task Group,* 11.

21. Barbara Anderson, "Web Lists or OPACs: Can We Have Our Cake and Eat It, Too?" *Library Computing* 18, no. 4 (1999): 312–16.

22. Ibid.

23. Rebecca Rollins, "Creating MARC Records from E-journal Title Lists," *IUG 2000 Conference Proceedings* (2000), www.innopacusers.org/iug2000/proceedings/j1.html (accessed Oct. 19, 2005).

24. Ibid.

25. Yiu-On Li and Shirley W. Leung, "Computer Cataloging of Electronic Journals in Unstable Aggregator Databases," *Library Resources & Technical Services* 45, no. 4 (2001): 198–211.

26. David Banush, Martin Kurth, and Jean Pajerek, "Rehabilitating Killer

Serials: An Automated Strategy for Maintaining E-Journal Metadata," *Library Resources & Technical Services* 49, no. 3 (2005): 190–203.

27. Siew-Phek T. Su, Yu Long, and Daniel E. Cromwell, "E2M: Automatic Generation of MARC-Formatted Metadata by Crawling E-Publications," *Information Technology and Libraries* 21, no. 4 (2002): 171–80.

28. Jimmie Lundgren, e-mail message to Iris Wolley, Dec. 21, 2005.

29. The Library of Congress Bibliographic Enrichment Advisory Team, *Web Cataloging Assistant* (2005). www.loc.gov/catdir/beat/webcat.html (accessed Feb. 6, 2006).

30. Ibid.

31. PCC Standing Committee on Automation Monograph Aggregator Task Group, *FREVR Final Report.*

32. Columbia University Libraries Digital Program, *Columbia HILCC: A Hierarchical Interface to LC Classification* (2002–). www.columbia.edu/cu/libraries/inside/projects/metadata/hilcc (accessed Aug. 24, 2006).

33. Library of Congress Processing Rule Analysis Group, *Cataloging Directorate Strategic Plan, Goal 4, Group 2: Processing Rule Analysis Report,* Dec. 9, 2003 version, rev. (Jan. 2004). www.loc.gov/catdir/stratplan/goal4wg2report.pdf (accessed Mar. 30, 2006).

34. Tom Delsey, *Defining an "Access Level" MARC/AACR Catalog Record: Project Report* (n.p.: Information Systems Support, 2004). www.loc.gov/catdir/access/report_final.pdf (accessed Jan. 25, 2006).

35. David Reser, *Access Level for Remote Access Electronic Resources: Results of the Access Level Test* (2005). www.loc.gov/catdir/access/access_test_public.ppt (accessed Jan. 25, 2006).

36. Library of Congress, Cataloging Policy and Support Office, *Appendix B: "Access Level" MARC/AACR Catalog Record: Mandatory Data Elements,* rev. Mar. 2005. www.loc.gov/catdir/access/mandatory_data_elements_test.pdf (accessed Aug. 23, 2006); Library of Congress, Cataloging Policy and Support Office, *Appendix C:*

"Access Level" MARC/AACR Catalog Record: Draft Cataloging Guidelines, rev. Mar. 2005. www.loc.gov/catdir/access/cataloging%20guidelines_test.pdf (accessed Aug. 23, 2006).

37. Library of Congress, Cataloging Policy and Support Office, *Appendix C: "Access Level" MARC/AACR Catalog Record.*

38. Access Level Record for Serials Working Group, *Access Level Record for Serials: Working Group Final Report* (July 24, 2006). www.loc.gov/acq/conser/pdf/alr/printer-version.pdf (accessed Aug. 23, 2006).

39. David Reser, *Defining an Access Level Record for Remote Access Electronic Resources* (2005). www.loc.gov/catdir/access/ala_crcc.ppt (accessed Aug. 23, 2006).

40. The guidelines for selection of free electronic resources are available upon request from the authors.

Appendix A. LC/CUL Comparison of Variable Data Fields and Their Use

Columbia University Libraries	Library of Congress
020 International standard book number	
$a, $z	$a, $z
022 International standard serial number	
$a, $y, $z	$a, $y, $z
024 Other standard identifier	
Not used	$a, $z
028 Publisher number	
Not used	$a
040 Cataloging source	
$a, $c, $d	$a, $c, $d
042 Authentication source	
Not used	$a Required
050 LC Classification	
Used	Used
240 Uniform Title	
If information is readily available. Use following appropriate LCRIs	Use following appropriate LCRIs
245 Title and Statement of Responsibility	
$a, $h, $b, $n, $p—do not transcribe other title information ($b) unless it provides needed information about the resource	$a, $h, $b, $n, $p—do not transcribe other title information ($b) unless it provides needed information about the resource
246 Varying form of title	
$a, $n, $p; first indicator = 1; second indicator = 3 only	$a, $n, $p
247 Former title or title variations	
$a, $n, $p	$a, $n, $p
250 Edition Statement	
$a	$a

4XX/8XX Series statement/added entry title

If clear that the resource forms part of a series, check appropriate authority files. If series is not under control, create an authority record.

If clear that the resource forms part of a series, and that series is one which LC does or would trace, create a series added entry, including volume/sequential designation as appropriate.

500 Viewed on note

Used

Not used

506 Restrictions on access note

$a—use only if a free subscription is required for access or for component parts of paid resources

$a, $b, $d, $e Use for notes from recommender/selector pertaining to restrictions on access and use imposed by a license or agreement through which the resource was acquired.

520 Summary, etc.

$a

$a

521 Target audience note

Not used.

$a Optional.

538 System details note

$a Used only if resource is not available via the World Wide Web.

$a Used only if resource is not available via the World Wide Web.

540 Terms governing use and reproductions

Not used.

$a, $b, $c, $d

773 Host item

$a, $t

$a, $t

780/785 Preceding/Succeeding entry

Not used.

$a, $t

856 Electronic location and access

$u, $z

$u, $3

Appendix B. Selector Survey

1. Are you using the new Internet Resources Cataloging Request (IRCR) Form for submitting cataloging requests for free electronic resources or component parts of subscription databases? (if no, please explain)

2. Does this new form and the electronic cataloging process impact your work? If so, how?

3. a. Is the IRCR form easy to locate on SWIFT? If not, where would you expect to find it?
   b. Is the IRCR form easy to use? If not, how could it be improved?

4. You are now able to track your submissions by using keyword searches and your UNI. How important is this to you?

5. How important is it to you to be able to contribute cataloging data such as keywords or summaries as part of the cataloging process?

6. The resulting bibliographic records are less full than those for RTIs.

   a. In your opinion, is there any important bibliographic information missing?
   b. How satisfied are you with this new record model from a selector and from a reference point of view?

7. How satisfied are you with the turn-around time of 1 to 3 working days?

8. General comments?