
Casting the Net

Caplan, Priscilla. "Cataloging Internet Resources." The Public-Access Computer Systems Review 4, no. 2 (1993): 61-66. To retrieve this file, send the following e-mail message to LISTSERV@UHUPVM1 or LISTSERV@UHUPVM1.UH.EDU: GET CAPLAN PRV4N2 F=MAIL.

Let Archie Do It?

How do we accommodate networked electronic information when our cataloging rules are designed to describe physical items owned by and residing in libraries? How do we provide access to that information? Do we let Archie do it instead? Questions like these must be addressed before we can move into the future and provide our patrons with information the way they are coming to expect it. It isn't sufficient that we simply debate these issues at conferences and write about them in the literature. Action is needed, and well-established rules and practices must be changed. All of that is easy to agree with, but deciding how to change established rules and practices is another matter, not to mention actually revising them.

What Are Online Information Resources?

MARBI is an ALA committee that advises the Library of Congress on additions and changes to the USMARC formats. Usually, MARBI deals with issues like where to record the International Standard Music Number and whether to add new coded values for Betacam videocassettes to the Physical Description Fixed Field. Last winter, however, a new proposal generated quite a bit of attention. Proposal 93-4 recommended changes to the bibliographic format to accommodate electronic data resources such as e-journals and documents available over the Internet.

Proposal 93-4 can trace its roots back to the summer of 1991 when MARBI Discussion Paper 49 proposed a set of data elements that might be useful in describing online information resources. Discussion of that paper, however, revealed some uncertainty about exactly what was being described. Was "online-ness" the salient quality, and if so, what did it mean to be online? Was it network accessibility? The property of being in electronic form?

Over the next year, some progress was made in sorting these things out. It became clear, for example, that remote access was the defining and unifying quality of these materials--the fact that they could not be held in the hand, physically described, pointed to on a shelf, or checked out to patrons. It was also agreed that for the sake of simplicity the universe of remotely accessed entities could be divided into two categories: (1) data resources (e.g., software, text and data files, and bibliographic databases) and (2) systems or services (e.g., campus-wide

information systems, library catalog systems, and bulletin boards). A rough but intuitive analog might be those things that one could FTP and those to which one could Telnet. Since electronic data resources more closely resemble what libraries are accustomed to cataloging than online systems do, MARBI decided in the winter of 1992 to concentrate on these first.

Joint Cataloging Project

Meanwhile, OCLC's Office of Research had received a grant from the Department of Education to investigate the nature of electronic information available over the Internet. OCLC's project staff had already collected and categorized more than 1,500 files. In the spring of 1992, representatives from the OCLC Internet Resources Project, MARBI, the Library of Congress, and the Online Audiovisual Catalogers teamed up for an experiment in cataloging electronic data resources. The group started with the hope that the existing USMARC computer files format could be used for remote data resources without too many modifications. This may sound simple, but for historical reasons the computer files format is surprisingly limited in its ability to describe computer files. It was originally designed with only a single type of file in mind--statistical data sets like Harris survey responses or the census. Later it was expanded to handle the microcomputer software that libraries had begun to collect. As such, it's like a house with only two rooms: a kitchen and a bedroom. OK until you want to take a bath.

Anyway, the project group took a sample of 300 data resources (mostly documents), drafted some preliminary cataloging guidelines, and sent the samples and guidelines to 30 volunteer catalogers so that each resource would be cataloged independently by three different catalogers. The volunteers were instructed to use the USMARC computer files format and AACR2 cataloging rules as best they could, and to keep a log of their particular problems, questions, and suggestions. The results were then compared, analyzed, and used to indicate where people were confused and where the format or the rules were deficient. The end products were a revised, more extensive set of cataloging guidelines and some recommended changes to the computer files format in the form of MARBI Proposal 93-4.

Recommended Changes

The recommended changes to the format for descriptive purposes were not extensive, but they required a modification to the cataloging rules. An existing MARC field called "File Characteristics" (256) is governed by AACR2 Chapter 9, which specifies that one of three terms must be used: "computer data," "computer program(s)," or "computer data and program(s)" (kitchen, bedroom, kitchen and bedroom). Proposal 93-4 recommended extending the set of allowable terms to include such descriptors as "electronic document," "electronic journal," "bibliographic database," "graphic," and "computer sounds." (A parallel set of coded values was defined in a fixed field data element to allow retrieval or reporting by these same concepts.) The rationale was simply to give brief, clear, descriptive information to the library patron, who might not intuitively think of an e-journal, for example, as "computer data."

Alas, expansion of the "File Characteristics" field was not approved by MARBI, on the grounds that the cataloging change must precede the format change. The issue was referred to CC:DA, another ALA committee that stands in very roughly the same relation to the cataloging rules as MARBI does to the USMARC format. As far as I know, CC:DA has not yet pronounced on this issue. Meanwhile, "computer data" it is.

+ Page 64 +

The biggest change proposed in Proposal 93-4 was not for the purpose of description but rather of location. In effect, it was decided that FTP sites, list servers, and the like constituted electronic locations that conceptually parallel physical ones. The paper form of a document might be on a shelf in a library, while a bitmapped form might be available from a file server on the Internet. A new field was invented for "Electronic Location and Access" (856), including data elements for type of access (e.g., e-mail, FTP, and Telnet), host name, path name, file name, and similar information necessary to access or retrieve a data resource over the network. Although much more radical an idea than the expanded list of file characteristics, this recommendation was independent of cataloging rules and so passed in slightly modified form at the January 1993 MARBI meeting. The "Electronic Location and Access" field is now formally part of the USMARC format.

Points to Ponder

At this point, it might be a good idea to pause and ask ourselves some questions. First, does it make sense for libraries to be cataloging Internet materials to begin with? Even if it does, is MARC the way to go about it? In fact, a vast number of data resources are available via the Internet, most of them uncontrolled, unverified, and of limited or ephemeral interest. (PACS-L readers may be reminded of the recent flap over an incomplete version of the Periodic Table.) Libraries are likely to have interest in only a small subset of this universe. For this subset, however, network access may actually be used to replace or supplement library ownership and physical access. Certainly, libraries will want these materials fully described. Similarly, such description or cataloging should be available in the same online catalog systems as the rest of the libraries' holdings, which implies that the records should be in the same format and follow the same rules as other bibliographic data.

+ Page 65 +

Won't network tools like Archie and Gopher supersede library catalogs for electronic data resources? These are wonderful tools, but they do have limitations we wouldn't tolerate for more traditional library materials. To quote another MARBI discussion paper (No. 69, April 30, 1993):

> Many do not give you any indication of which servers they actually searched and which were unavailable for one reason or another. They do not discriminate between various versions of data in terms of usefulness or completeness. They are poor at locating known items, as opposed to possibly relevant things. In addition, the subject analysis available in USMARC records is lacking in these other tools.

> . . . Such tools could complement rather than replace USMARC records as a source for locating online objects.

But electronic addresses change often as documents move from server to server and from format to format. Does it make sense to actually imbed location information in the descriptive record itself? Well, probably not. In fact, the Internet Engineering Task Force is working on a much more efficient scheme. A Universal Resource Identifier (URI)--much like the ISBN--would be assigned to each object by the originating agency. A Universal Resource Locator (URL), similar in concept to the Electronic Location and Access field, would identify a location. Only URIs would be imbedded in the bibliographic description, and computers would associate the URI with one or more URLs in much the same way an Internet host name (HARVARDA.HARVARD.EDU) is associated with its IP address (128.103.60.11) by the name server system. However, someone needs to do all this; an infrastructure needs to be developed and responsible agencies in agreement on responsibilities and procedures. Once this mechanism is in place, we can decide what to do next with the Electronic Location and Access field. Meanwhile, it allows us to begin building records and testing the feasibility of catalog access to electronic data resources.

Finally, can we really separate data resources from systems/services? Is there a distinction between a database and the retrieval system required to access it? Probably not. Although a useful distinction to get the project going, the line between these concepts was fuzzy to begin with and gets fuzzier the more you think about it.

Next Steps

The next step is clearly to try to accommodate online systems and services in USMARC as well. Some of the relevant data elements such as hours of service and cost for use don't currently exist in the bibliographic formats, but they are defined in the new USMARC Community Information Format. We may end up with a hybrid that seems "bibliographic" in some respects and like a program or agency in others. Discussion Paper no. 69, "Accommodating Online Systems and Services in USMARC," addresses these issues. It will come up for discussion when MARBI meets at the ALA annual conference in New Orleans. You can request the paper by sending this e-mail message: GET DP69 DOC to LISTSERV@MAINE.MAINE.EDU.

Or, as the Electronic Location and Access field would have it:

```
856 0  $a maine.maine.edu $f dp69 doc $h listserv $i get
```

About the Author

Priscilla Caplan, Head, Analysis and Programming Division, Office for Information Systems, Harvard University Library. Internet: COTTON@HARVARDA.HARVARD.EDU.

The Public-Access Computer Systems Review is an electronic journal that is distributed on BITNET, Internet, and other computer networks. There is no subscription fee.

To subscribe, send an e-mail message to LISTSERV@UHUPVM1 (BITNET) or LISTSERV@UHUPVM1.UH.EDU (Internet) that says: SUBSCRIBE PACS-P First Name Last Name. PACS-P subscribers also receive two electronic newsletters: Current Cites and Public-Access Computer Systems News.
