SPEC

Flyer 249

Cataloging of Resources Digitized for Preservation September 1999

Introduction

The majority of current digital preservation efforts in research libraries still operate as pilot projects. Not surprisingly, many of the pioneering projects focused on determining the costs of scanning, the organization of information, design of the user interface, and development of the technical infrastructure for maintaining digital resources over time. Far less attention, outside of a few notable exceptions, was paid to issues related to describing or cataloging of digitized resources. With the advent of the World Wide Web and the demand for effective user access over the Internet, this situation is gradually changing.

As libraries undertake more digitization projects and as plans for large-scale imaging projects move forward, effective access strategies will play a key role in facilitating resource discovery. The mainstay of libraries' cataloging practices has been the provision of individual resource descriptions, that is, online bibliographic records in the library's OPAC and major bibliographic utilities. A central component of these efforts has been adherence to AACR2 (Anglo-American Cataloging Rules) and MARC (Machine-Readable Cataloging) standards, use of Library of Congress subject headings, and authority control of name and subject entries.

User demand for web access, rapid changes in information technology, and the high costs of MARC cataloging, however, are fueling a transformation in the methods for access. Not only are there several changes underway in the United States in AACR2 and MARC standards, but also on the international front in such efforts as the International Standard Bibliographic Description for Electronic Resources. Most importantly, libraries are pursuing entirely new approaches to facilitate access to digitally reproduced resources. Major efforts are underway to provide user access to these resources on the World Wide Web. The spectrum of options ranges from full-text encoding to provision of metadata modeled after the Dublin Core to collection-level MARC records and encoded archival finding aids.

This SPEC survey focused on current bibliographic control strategies for materials digitized for preservation. The survey was intentionally limited to electronic access to materials that were digitized from a library's own collections for preservation purposes. The survey did not attempt to deal with cataloging of remotely accessed electronic resources or resources that only exist in digital form.

Survey Results

Sixty-one of the 122 ARL member libraries (50%) responded to the survey. Only 19 respondents (31%) reported that items digitized for preservation purposes are cataloged, while 26 reported no cataloging activity. Sixteen institutions indicated that planning is underway. Comparing these results with the responses to a survey conducted in 1996, *Digitizing Technologies for Preservation* (SPEC Kit 214, March 1996) institutional practice has not improved significantly over the past three years. In the 1996 survey, 28% of the respondents indicated that MARC records are created for digitized materials. In both surveys, the focus was on the use of digitization for preservation purposes.

Collectively, the respondents to the new survey reported that 255,467 paper-based items were digitized in 1997–98. Projects ranged from a few items to 130,000, with a median figure of 1,100 items. Although the survey did not ask about titles or types of materials digitized, the responses indicate a striking variety of newly digitized resources. Among these collections are photographs, posters, manuscripts, papyri, and archives.

Cataloging Policies. Nearly all of the institutions that currently catalog items digitized for preservation purposes reported that MARC and AACR2 standards are followed. But institutional cataloging practices vary considerably in the specific cataloging technique, depending upon local constraints, compromises between cost and availability of existing records, and the ease of creation and maintenance. Twelve institutions (57%) reported that they amend an existing catalog record for the nondigital original to provide information about the digitized resource. Nine libraries (43%) create a separate machine-readable record for digitized monographs and serials. Moreover, policies are often situational and may vary based on the specific requirements of a project or on system constraints.

Efforts are underway to standardize practice for cataloging electronic resources, most notably through the Library of Congress's *Draft Interim Guidelines for Cataloging Electronic Resources* (available at <http://lcweb.loc.gov/catdir/cpso/elec_res.html>). The Library of Congress (LC) policy uses the single-record approach for digitization of existing LC collections and "one-to-one" electronic reproductions and the multiple-record approach for compilations and electronic anthologies.

One challenge looms large: providing access to resources such as manuscripts, posters, or photographs that libraries traditionally have not described individually. Several libraries reported using collection-level records for digitized compilations, while others rely on metadata in the digitized files only.

Managerial Responsibility. Cataloging of electronic resources, like other kinds of resources, is a highly technical matter, and in 28 responding libraries (85%) the responsibility rests with the cataloging unit. In five libraries (15%), the unit responsible for creating the electronic resources also manages the description of the digital files.

Networked-Based Access to Digitized Resources. The most striking aspect of the responses is the extent of experimentation and the range of approaches in the use of web-based access strategies. The various access mechanisms represent widely varying costs and implications for future technological improvements. The present state of affairs makes it difficult, however, to determine whether an item has been already digitized, because there are so few standardized registers or catalogs of the growing array of library digitization projects. One exception is the Association of Research Libraries' Digital Initiatives Database <http://www.arl.org/did/>. This is a web-based registry for descriptions of digital initiatives in or involving libraries. The goal of the effort is to capture basic information for a wide range of digital initiatives. The database is a collaboration between the University of Illinois at Chicago and ARL.

Like the variety shown in network-based access strategies, there also exist a multiplicity of options for describing digitized resources. The Dublin Core Metadata Initiative offers a lower-cost alternative to full MARC cataloging. Dublin Core records encompass information about the technical characteristics of digital files, their location, and a summary of their content. Several libraries reported using Dublin Core descriptions that are embedded in the electronic resource. Other libraries reported using descriptors that are modeled after the Dublin Core. OCLC's Cooperative Online Resource Catalog (CORC) project is pioneering efforts "to build the next-generation catalog of Web resources." Already more than 30 ARL libraries are cataloging digital resources as part of the CORC project, and several respondents indicated that they plan to participate in the project. Still another approach is to create separate access databases, such as the Digital Initiatives Database, to help users in discovering and accessing digitized resources. This approach is significantly enhanced by the capacity of new OPAC records to link directly to websites. Many other possibilities are on the horizon. While the responses show a range of options for describing digitized resources, it is evident that no single approach has been widely adopted yet.

Conclusion

Many factors are contributing to the transformation of traditional approaches in making research resources known to users. Increasingly, these approaches focus on making resources available through the web. To make matters worse, not only are these strategies for accessing digitized resources changing rapidly, they also exist now in a wide variety of choices. Since the pace of change seems unimpeded by the lack of standards and uncertainty about future technological developments, libraries engaged in digitizing for preservation must determine the best approach to digital capture as well as the most effective access strategies in a dynamic environment.

This SPEC Flyer and Kit were prepared by Jutta Reed-Scott, ARL Consultant.