ORGANIZING AND RETRIEVING IMAGES

Capturing and storing images improves access only if the users can easily search and identify them. Although full-text files can be searched by keyword on all or part of the text, software that searches graphical elements in digital image files is still in the early stages of development. Therefore, you must create a catalog or index.

Markup

Even before images can be cataloged or indexed, markup should be undertaken. Markup is a form of text added to an image to transmit information about both the source document and the image. Markup is not unique to imaging. Printers have been using it for centuries to instruct a typesetter about layout, fonts, italicization, capitalization, and so on.

In the electronic environment, there is a generalized markup language. It is not tied to any hardware or software, so documents using a generalized markup language can be migrated from one hardware-software environment to another. Another characteristic of a generalized markup language is that it allows for encoding based on the content of the work.

HTML

The most widely used markup language for images is the HyperText Markup Language (HTML). Developed for publishing hypertext on the Web, it has appeared in four main versions, with the World Wide Web Consortium (W3C) recommending HTML 4.0 as the markup language of choice. HTML is a derivative of SGML (Standard Generalized Markup Language) and is more useful than it because it is relatively simple to learn and is directly accessible to most browsers.

For those who cannot spare the time to learn HTML, there are text creation programs.

Markup is also a service included by a service bureau that scans images.

File Names

File names should be established for all images during markup. The file name should uniquely identify the image. The file name can be a name, call number, or accession number, or a combination of these. The file name can also be imbedded in the digital image using image-editing software such as Adobe Photoshop. Thus, in addition to a bibliographic record that links to the image, the file name provides a unique identifier.

Cataloging

There are three cataloging options for images: MARC, Dublin Core, and EAD.

MARC Format

The most readily available method of organizing and retrieving images in libraries and archives is the MARC format and a link from its 856 field to the digital

image. This approach not only provides access to the images per an internationally adopted standard but integrates access to images with other materials in the organization's collections. Adherence to the standard lets users exchange records among libraries or combine records in a union database, either physical or virtual.

Although the cataloging of images of books, periodical articles, pamphlets, and other published materials is straightforward, libraries do sometimes experience problems with the cataloging of images of photographs, maps, and other materials that lack title pages. The following guidelines are useful in such cases:

1xx Main Entry

Use the name of the person or corporate entity that was responsible for the creation of the item. That could be a photographer, artist, mapmaker, newspaper, map publisher, and so on.

245$a Title

If there is none on the image, use the caption in the associated text, if any. If there is none, create a brief descriptive title and enclose it in square brackets.

245$b Additional description

Add additional descriptive text.

245$c Statement of Responsibility

Record a single statement of responsibility; record attributions in square brackets.

246 Series Title

Enter the name of the series of which the source document is a part.

260 Publication

Include name of the publication and date, if known.

300 Physical Description

Gives the general material designation in $a, the other physical description in $b, and the dimensions in $c.

500 Notes

Record the source of the title, physical damage or alteration, inscriptions, stamps, or other markings.

506 Terms of Access, Use, Reproduction

Itemize any restrictions.

520 Subject Description

Include a brief description of what is depicted.

541 Immediate Source of Acquisition

Record how the material was acquired, from whom, date, and price, donation, or loan status.

600 Subjects

600/610—Limit to well known people or institutions.

650—Chose no more than three or four Library of Congress (or other) subject headings.

651—Enter geographical heading if appropriate.

655—Enter a specific heading describing the recording process of the source document.

7xx Added Entries

700/710—Enter donor name with $edonor.

856 Unique Identifier

Enter the unique identifier that links the record to the image.

If the source document and image are part of a folder, you can use a 852 tag and transcribe the heading on the folder.

The drawback to using the MARC format is that it is not flexible, and it changes only slowly. For example, MARC does not accommodate the arrangement of related images in multiple groups within a collection and many important details about the history of an image have to be included in a general notes field.

The two other major formats do a better job of accommodating information about the digital image, including:

- Information required to view the image, including format and compression scheme
- Information about the quality and accuracy of the image
- Information about the scanning process, including resolution, date of scan, a record of image manipulation, and so on
- Rights and reproduction information
- Information about the location of the source document

Dublin Core

Dublin Core is a set of 15 core elements that can be used to describe electronic resources: title, creator, subject, description, publisher, contributor, date, type, format, identifier, source, language, relation, coverage, and rights.

Dublin Core was originally intended for resources hosted on the Web so it supports HTML and XML. The elements were designed for flexibility and can be mapped to MARC records elements for conversion, or for the creation of metafiles that include both Dublin Core and MARC records.

Dublin Core is gradually displacing MARC as the approach for organizing images in libraries, especially in libraries with huge image files or that seek to provide access primarily via the Web. Those libraries familiar with Dublin Core are of the opinion that Dublin Core is more economical than MARC, but no comparative cost data appears to be available.

More information about Dublin Core can be found at **http:purl.org/dc**.

Dublin Core was not yet a standard as of December 2000, but the National Information Standards Organization (NISO) had completed balloting on the standard and expected to publish Dublin Core as a standard in the first half of 2001.

EAD (Encoded Archival Description)

The Society of American Archivists and the Library of Congress' Standards Office have developed a metadata cataloging standard for archival collections that use SGML (Standard Generalized Markup Language. XML can be used as an option. The standard makes provision for extensive scope and content notes, historical background, administrative information, and series description.

EAD provides for both collection-level finding aids and MARC-like individual cataloging records called SIM (Single Item Metadata) for individual items in the collection.

EAD must be viewed with an SGML viewer. Although such viewers are readily available by downloading from the Web without charge, they do not support printing. XML viewers are still not widely available. This drawback can be overcome by using a relational database that is ODBC-compliant (Open Database Connectivity) such as Oracle to create the record. Then you can output the record in any appropriate Web language: HTML, SGML, or XML.

An overview of the EAD structure is available on the Library of Congress' Web site at **http://lcweb.loc.gov/ead.html**.

EAD is not well known to librarians, but is widely used by archivists.

Indexing

Cataloging is a time-consuming and costly process. Many libraries and archives have chosen to forego cataloging and have purchased an indexing tool designed for imaging applications in business and government. The most widely used is Kodak's QuickSolve, a software package originally developed for the law enforcement community to manage digital images and other caseload data. It provides access to thumbnails of images by keywords and many other customized fields, and links to the service or screen images. QuickSolve can be used on a network.

For more information, or to download a trial of the software, go to **www.kodak.com/global/en/professional/hub/lawproducts/software/quick.shtml/**

Clearinghouse of Image Databases

Libraries and archives that develop image databases should contribute information about their databases to the Clearinghouses of Image Databases operated by the University of Arizona Library ***(www.library.arizona.edu/images/image_projects.html)****. The Clearinghouse's databases can be searched by keyword to identify related materials to which a library or archive may want to provide links from its patron access catalog.*