

Historical collections for the National
Digital Library: Lessons and challenges at
the Library of Congress.

□

Historical Collections for the National Digital Library: Lessons and Challenges at the Library of Congress

Continued from [part 1 of this article](#)

Caroline R. Arms
Information Technology Services & National Digital Library Program
Library of Congress
caar@loc.gov

D-Lib Magazine, May 1996

ISSN 1082-9873

□

<http://lcweb2.loc.gov/ammem/ammemhome.html>

Contents

- [Introduction](#)
- [Lessons and Challenges -- part 2](#)
 - [Guiding users through the growing archive is a challenge](#)
 - [Finding the right words can be difficult for users](#)
 - [The computer screen has limitations](#)
 - [The digital archive must be managed and preserved](#)
 - [Even for historical materials, access restrictions must be considered](#)
- [What Does the Future Hold?](#)
- [Footnotes, References, and URLs](#) (in a separate document)
- [Indexing the historical collections](#) (sidebar in a separate document)

Introduction

This is the second of a two-part story on the National Digital Library Program (NDLP), the project at the Library of Congress to assemble and make publicly available an archive of millions of digitized reproductions of primary source material for the study of American history and culture. These historical collections, collectively known as American Memory, form one focus of the Library's effort to provide widespread access to its resources and services by taking advantage of the potential of networked access to information in digital form. Among the other projects are [THOMAS](#), for public access to legislative information, and a prototype system ([CORDS](#)) for accepting materials in digital form for copyright registration and deposit.

Some key points discussed in the earlier part of the article are important as context for the second part.

- The Library of Congress is assembling its historical digital archive by collections, each collection being a grouping of digital reproductions that form an integral whole. This approach is consistent with archival practice, which provides context for individual documents by arranging them within a collection of related materials.

[For examples of some of the current collections, click on one of the images in the collage at the top of the article. Each image links to the "home" page for the collection to which it belongs. Alternatively, browse the [list of collections currently available](#).]

- Archival collections have often not been described by individual records for each item, but by documents that list items or groups of items, and guide the user through the intellectual and physical organization of the material. The preparation of such "finding aids" is much less demanding of staff time than the provision of individual bibliographic descriptions through catalog records for each item. Designs for a digital library architecture must recognize this aspect of archival practice.

Since the first part of the article was published in mid-April, the Library has begun to encode some existing finding aids using the new draft standard for Encoded Archival Description (EAD). EAD is a document type definition (DTD) for the Standard Generalized Markup Language (SGML). The finding aids that are being marked up initially describe collections that have not been digitized, but the standard supports direct links to digital items. Readers with an SGML viewer can see the first [sample finding aids](#).

- Each collection, and each item within a collection, has been given a logical name that will be the basis for a persistent, location-independent Uniform Resource Name (URN) when such names are supported on the Internet.

Lessons Learned and Challenges Posed -- part 2

Guiding users through the growing archive is a challenge

Providing navigational tools to support effective access to the resources in the growing digital archive will be a challenge. The resources must be presented for a diverse range of patrons: for teachers, students, and the occasional user as well as for the librarian and the scholarly researcher. Some users will have limited time and a very specific task at hand; they will benefit from the ability to search with precision, perhaps for a speech by a particular individual or pictures of hotels in Detroit. Others will have plenty of time and want to gain an overview of what is available in the entire archive or on a general topic before searching for anything in particular. They may prefer to read introductory presentations, take guided tours, browse lists of topical headings, or perform a general search and look at large numbers of individual items.

The American Memory archive of historical materials is being assembled collection by collection; these collections have different characteristics. Collections currently available over the World Wide Web consist of a single general type of material (text, image, audio, or movie), but collections in process incorporate several types. Later this year, two thematic anthologies comprising textual documents and images will be released, one on the American variety stage and one on the history and development of the environmental conservation movement. Another collection in process, covering the presidency of Calvin Coolidge and the transition to a consumer economy in the 1920s, includes all four types of material. Some digital collections will have individual bibliographic information (catalog records) for each item, with individually assigned subject headings. Others will be described primarily in a finding aid, a structured document with links to the items listed. The intellectual context and depth or breadth of focus for collections also varies. Some comprise a large proportion of the output of an individual creator. Some are personal selections of collectors or scholars. Others are collections of materials in a particular physical format with no single topical theme.

Initially, the only way to access American Memory was to start by browsing [a list of the collections](#). Each collection was described by a sentence and a list of broad subject topics. By early 1996, there was a full list, of about eight collections, and a subsidiary list for the four general types of material. By the time the latest collections were released in March 1996, the ability to [search across all the historical collections](#) simultaneously had been added. It is also possible to search across collections containing items of a certain type (text, image, audio, or movie). For a cross-collection search, the list of hits identifies the parent collection and the item type, as well as the item's title.

Each collection currently consists of digital reproductions, a group of linked HTML pages providing archival and intellectual context, and one or more access tools for exploring the collection. Varying with type of material and level of description available, the access tools within a collection are chosen from the

following:

Searching bibliographic records

Most of the current American Memory collections have a set of item-level bibliographic records which can be searched. The search engine in use is INQUERY, the text search toolkit from the [Center for Intelligent Information Retrieval](#). [For more detail on the indexing of the historical collections, [see sidebar](#).] A ranked list of "hits" is returned. INQUERY's flexible query commands have been used to list the hits for queries of two or more terms in four sublists:

1. items that contain the entered query as an exact phrase;
2. items that have all the search words within 20 words of each other;
3. items containing all the words, but not close to each other;
4. items with one or more of the entered words, but not all.

The title displayed for each hit acts as a link to a full bibliographic record display, which has a link to the item. For images, the link in the bibliographic display is a thumbnail version of the image. Each subject term displayed in a retrieved bibliographic record is also a link that invokes a search for items on the same topic. Names of authors, photographers, and other creators are links to searches for other works by the same person or organization.

Many of the presentation details, including the division of the hit list into sections, have been modified after user feedback. Exact phrase matches are presented first, since analysis of queries entered on the THOMAS system over a 3-month period indicated that a large proportion of queries were single words or simple phrases [1]. An informal study in the Prints and Photographs reading room involved ten users performing two pre-selected image retrieval tasks using an earlier version of the American Memory interface. The searchers were encouraged to talk as they searched; reference librarians logged their comments and the difficulties they were observed to have. Several users were very concerned when searches returned many results that were not close matches to their query; two explicitly requested that the system indicate if there were no items that matched all the entered terms. The further division of the results list addresses this concern.

Browsing subject terms

For each collection with item-level bibliographic records, a browsable alphabetical list of all formal subject headings has been compiled. These subject headings typically include names of people and places, terms describing the genre and format of the original item, and topical descriptors. For an example, see the [subject listing for the collection of over 25,000 photographs from the Detroit Publishing Company](#).

Searching full text

For text collections, users can search the full text of documents. INQUERY is used for indexing and retrieval and the ranked list of hits is subdivided in the same way as for bibliographic searches. Each title in the result list acts as a link to an HTML presentation of the corresponding document. Full-text search is the primary access tool for [Life Histories gathered during the Federal Writers' Project](#), a project of the Works Progress Administration (WPA) in the 1930s, since there are no individual bibliographic records. For text collections with bibliographic records or SGML versions of the documents, there are corresponding links from the HTML version of the full text.

Finding aids

In the future, browsable finding aid documents will probably be used as the primary access aid for many collections where item-level cataloging is not feasible. Currently, two very simple finding aids are in use: for the WPA Life Histories, presented through a hierarchical list organized by region and state; and for the four notebooks of Walt Whitman that were recovered in 1994. Once the Library begins to use the Encoded Archival Description format for finding aids to digitized collections, finding aids will also be indexed for searching.

Another access path is provided through the Library's main on-line catalog, which has collection-level descriptive records. These catalog records indicate that the collection is available on the World Wide Web. Each record contains a pointer (in field 856 of the MARC format) to the collection by its assigned logical name, ready for future implementation of direct links. Because records from the Library of Congress catalog are distributed widely as the basis for other libraries' catalogs, the Library is hesitant to add location-dependent identifiers, such as URLs that will change over the years, to its catalog records.

The combination of navigational tools supports a variety of approaches to identifying collections and browsing or searching for individual items. However, it is clear that as the archive grows, new approaches will be needed. Today it is possible to scan the list of collections in a few minutes; an hour or two gives time to dip into each collection and form an impression of its content. But collection identification through browsing a list with brief descriptions, although appealing to users, will not scale up as the number of collections increases and digitized historical collections at many institutions are incorporated into a truly national, distributed digital library. Better methods for quickly identifying the collections that are most likely to be of interest will be needed. The User Interface Team from the National Digital Library Program is currently working with the [Human Computer Interaction Laboratory at the University of Maryland](#) to explore new approaches for the user interface.

The challenge of guiding the selection of collections has at least two aspects:

determining the characteristics that best describe and distinguish the collections; and finding ways to present those characteristics in a fashion that lets the user select relevant collections easily. As an alternative or supplement to the traditional text-based searching of catalog records, the User Interface Team plans to explore the use of direct manipulation and rapid feedback during the selection of collections through browsing and filtering operations. To support these operations in a prototype interface, the team will build an independent set of descriptive (but non-MARC) records for the collections using a format that is easily modified and indexed by INQUERY, but not intended for viewing by users. Among the key questions for this part of the interface and its supporting database are how to represent time and place and how broad or detailed subject topics should be. After the experimental stage, during which the choice and format of attributes will be refined, the Library expects to integrate the content of these records into any standard architecture for collection-level metadata that emerges through broader digital library efforts, such as the series of Metadata Workshops sponsored by OCLC in conjunction with other organizations, the development of a Z39.50 profile for access to digital collections, or cooperative ventures through library consortia such as the National Digital Library Federation.

Guiding users to relevant collections is not the only navigational challenge for the user interface. When most items in the digital archive have individual bibliographic records, a search across all items is straightforward. The level of description is roughly comparable and catalogers take considerable effort to be consistent (although consistency is more common within collections than across different collections). The granularity of the chunks of text indexed are also comparable. However, the historical collections will soon provide a more complex structure. Some items will be described only through section headings and very brief descriptions in a structured finding aid. Textual items will be fully indexed by the words they contain. It is easy to pour all this text into an indexing engine. It is more difficult to structure search options and present results in a way that is both efficient and comprehensible to users.

Another dimension of the challenge in building access tools is that a single interface is unlikely to satisfy the entire range of users. Users will be engaged in very different tasks, from the detailed study of a period in the history of photography to picking one photograph to illustrate a high-school paper on the Civil War. As part of its educational outreach, the National Digital Library Program has already developed materials and tips aimed particularly at teachers. The initial search interface for American Memory's historical collections was made as simple as possible. Experienced searchers have requested the ability to formulate more precise searches. More complex query forms involving Boolean logic and searching for terms in specific fields are under test in other projects using INQUERY at the Library, for example, in the *Digital One-Box* that provides access to image collections in the Prints and Photographs reading room. It is likely that an "expert" search query form will be developed for the historical collections.

Finding the right words can be difficult for users

The ability to specify queries more precisely will address some searching problems, but mis-matches in terminology also prevent users finding relevant items. Researchers making extensive use of the Library's traditional collections soon learn the subject headings that relate to their field of interest. However, casual users do not naturally use the terms from controlled vocabularies. Topical subject headings for bibliographic records are chosen from controlled vocabularies, such as the Library of Congress Subject Headings (LCSH) and the [Thesaurus for Graphic Materials](#) (TGM). A search for photographs using the term "state house" retrieves 50 pictures for which the exact phrase occurs in the bibliographic description. However, the TGM specifies the synonym "capitols" as the appropriate subject index term and a search on "capitols" retrieves 256 photographs (only a few of which are of the U.S. Capitol).

In many contexts for information retrieval (for example, scientists accessing current literature in their field), the primary users can be assumed to be familiar with the terms in the documents themselves. For historical texts, that is not necessarily the case, especially for occasional users. Usage changes over time. "Teacher training" is unlikely to appear in a nineteenth century document, but teachers were certainly trained and the collections currently mounted have much to say about this topic, particularly for women (in the [National American Woman Suffrage Association collection](#)) and Afro-Americans (in the collection of [African-American Pamphlets](#)). A recent search found the phrase "normal school" (the type of school at which teachers were trained) occurring 66 times in the historical collections, whereas "teacher training" appears only twice.

Finding the right level of specificity for search terms is also a problem for users. Teachers might be looking for materials to support their treatment of broad concepts and movements covered in the K-12 curriculum, such as "mass transportation," "leisure activities," "urbanization," or "economy of the South." Cataloging guidelines usually recommend the application of the most specific terms, with an emphasis on the concrete. Searching for photographs on "leisure" retrieves 10 examples of people relaxing doing nothing in particular. However, there are 36 photographs related to swimming, and 19 with the phrase "state fair" in their description. Primary textual materials that illustrate a topic may not contain the common general term for the topic. A pamphlet entitled "Illiteracy and its social, political and industrial effects" has a comparison of taxes and expenditure for education between northern and southern states. The pamphlet will be found by searching on words such as "tax", "wealth," "poverty," "income," and "capital," but it does not include the word "economy."

One potential approach is to integrate thesauri into the interface, by providing a browsable hierarchy of terms to assist selection, or by mapping commonly used terms into formal subject headings or to synonyms actually found in text documents. The Library will be exploring possibilities in this area in at least two

ways. Firstly, the approach used currently for automatically expanding a search query to search for plurals and other word variants can be extended to incorporate synonyms. Secondly, the Library is working with [Project Management Enterprises, Inc.](#), the developers of [LEXICO/2](#), a thesaurus management system used by several divisions in the Library to manage thesauri, such as the [Thesaurus for Graphical Materials](#) and the [Legislative Indexing Vocabulary](#).

The computer screen has limitations

The computer offers rapid access to an array of resources, but its screen is less convenient than a large table for studying images and than a book for skimming through indexes and pages of text to determine its relevance.

The images the Library has disseminated to date have been sized for display on today's computer screens, usually between 500 and 1200 pixels wide. For photographs and page images, these "reference" images are satisfactory for many uses, although many researchers and publishers will still want to see the original or take advantage of a high-resolution version at the cost of time to download (not currently supported). However, if maps and large posters are scaled to fit on a screen, too much detail may be lost for all but the most cursory scanning. The Library hopes that other organizations will develop approaches (and related viewers) allowing users to zoom in on selected areas with progressively higher resolution. A promising line of research is the application of wavelet transformations to a browsing model for large images at the [Alexandria Project at the University of California, Santa Barbara](#).

Visitors to the Prints and Photographs reading room can explore image collections by skimming quickly through a folder full of similar images. They can pick out those that catch their eye and lay them out on the table. This is particularly valuable for comparing similar images. One approach to providing equivalent functionality is to display many thumbnail images in a grid. Of the ten searchers observed performing two specified retrieval tasks using American Memory in the Prints and Photographs reading room, two suggested that a set of thumbnails (usually about 150 pixels wide) would be useful as a hit list, but another commented that she always had to expand each image to determine whether it was useful. Finding an appropriate balance between detail visible and response time may not be easy, since users will have differing tasks and priorities.

Although an increasing volume of primary textual materials is available online, there is no consensus on how best to present long documents in a way that supports convenient use. The challenge is to support rapid navigation without losing the sense of context provided by a physical book. Most of the books, papers, and pamphlets in the historical collections have been coded in SGML with embedded links to images of the original pages, illustrations, and tables. The SGML representation (using a document type definition based on the guidelines of the [Text-Encoding Initiative](#)) captures features of documents that provide

great potential for both convenient presentation and effective searching, but the potential has not yet been fully exploited.

American Memory resources must also be accessible to the widest possible audience, and convenient for use in the classroom. Today, users of MS Windows who are comfortable with the technology can download a free SGML viewer (Panorama Free). This free viewer can present a structured table of contents derived from chapter and section headings. Headings act as direct links to corresponding sections of a document. However, the internal search feature of Panorama Free does not take advantage of the structural information encoded in the SGML format. Nor does Panorama Free support printing. To provide access to a much greater range of the public, including users of other operating systems and proprietary browsers from an on-line service, and those without easy access to technical support, the Library has converted the documents into HTML. The book-length works are somewhat clumsy in their HTML form, but this format does allow any user to download or print the text of the entire work reasonably conveniently -- absolutely necessary for use in schools where equipment for access to the Internet is, at best, a scarce resource.

The problem of networked access to book-length works is not unique to the Library of Congress. As an increasing array of documents are available in SGML format, the Library hopes that a more powerful SGML viewer will become widely available or that WWW browsers will be extended to handle SGML. Meanwhile, the Library will be experimenting with other options, such as Adobe's PDF format, the use of frames within HTML, and the development of viewer modules using the Java programming language. Meanwhile, a variant approach using HTML will be taken for the [Country Studies/Area Handbook Program](#). These studies have been prepared by the Library's Federal Research Division for over 60 countries around the world.

The digital archive must be managed and preserved

Currently, contractors deliver files to the Library on CD-ROM. As for many other imaging projects, the highest-resolution images have been stored offline to save disk space. Compressed files for on-line delivery of images are loaded onto disk drives attached to the WWW servers. The CD-ROMs serve as an archival copy for the uncompressed versions. As the size of the archive grows, these CD-ROMS require formal inventory control. Ensuring the continuing usability of the archival files is also important since the accumulated effort involved in their creation will be valued in millions of dollars. The need to manage vast archives of computer files is not unique to libraries; solutions are appearing from the commercial sector.

To address the estimated need for 50 terabytes of managed storage by the year 2000, the Library will be installing a commercial hierarchical storage management system over the coming months. Storage management software will transfer files

automatically between high- performance disk drives and less expensive storage media from which retrieval will take longer. Typically, the location is based on time since last access, but more complex rules can be enforced. Since the highest resolution images will be accessed only occasionally, they will usually be resident on the slower medium. The allocation is transparent to applications accessing the files; "stub" files left in the logical file hierarchy point to the physical location of files that have been relegated to another "layer" of storage. For its second layer of storage, the Library has chosen high capacity magnetic tape cartridges under robotic control.

The tape unit ([IBM's 3494 Tape Library Dataserver](#)) has been installed and is already in use for regular backup and restore operations (associated with any computer system). Generation (and periodic regeneration) of archival copies of collections on tape cartridges is also possible for preservation of the digital materials. The hierarchical storage management software ([ADSM from IBM](#)) will be installed on individual computers over the next few months, in conjunction with system upgrades.

Direct access to the archival digital reproductions will also support other functions. Online access to high-resolution images could facilitate digital reproduction as an alternative to the Library's current service that provides users with photographic copies of graphical materials for a fee. The expectations of users in the Prints and Photographs reading room have already been raised. After identifying the images they need in a few minutes, some users express dismay that photoduplication requests may take several weeks. The National Digital Library Program also plans to disseminate materials in other forms, perhaps supplying source files to companies or organizations that will assemble subsets and tailor them to particular markets. Providing source files will be easier and more reliable with on-line access than by manual sorting through CD-ROMs and copying large numbers of files.

The hierarchical storage management system will address issues related to physical storage of files, but there is also a need for more formal management of the collections and objects themselves than can be provided by the Unix file system. The organization of collections is currently through the relationships between naming schemes and Unix directories. There is no automatic enforcement of naming conventions. If a digital item has several component files (such as a set of page images for a document), there is no way to treat that item reliably as a unit. One of the custodial divisions keeps track of directories on charts posted on an office wall. The tables that map logical names into physical locations during retrieval are in a simple ASCII file. If a collection must be moved (for example, as the archive grows and must be divided between servers), the charts and the file must be modified in concert. While the number of collections is small, this approach is feasible, but the Library recognizes that it will need enforcement of unique names, more automatic adjustment of all pointers when files are relocated, and a structure that recognizes key relationships between

items and their component files, and between collections and the items they comprise.

Since no adequate system has emerged from the commercial marketplace to support effective access to multimedia resources of this complexity, the Library has begun work with collaborators to design and build two prototype "repositories" as experiments in this area. A preliminary set of attributes or fields has been developed for use in both repositories, a set that is limited to metadata relating to the "physical" digital object, and its creation, modification, retrieval, and display. This "physical" metadata includes the logical name, access privileges, the digital format, key structural relationships (such as sequenced pages within a book, or segments of a poster digitized as several files), and administrative data such as the unit or person responsible for the item, date created, and number of times accessed. Primary intellectual access to both repositories will be through the related bibliographic records (in MARC format) or finding aids discussed earlier.

CNRI repository prototype

One prototype repository is under development with the Corporation for National Research Initiatives (CNRI). This repository will also be used for the [CORDS](#) project in the Copyright Office. The design uses the CNRI handle system to give location-independent names (URNs) to digital objects in the collections and is based on the [distributed framework proposed by Kahn and Wilensky \[2\]](#). The repository provides a secure environment from which the items in the collections and their metadata can be accessed together or separately. This permits negotiation over access terms and conditions for a digital object to be an integral part of every access, access which must be possible through many protocols, such as HTTP and Z39.50. The repository will allow for an essentially limitless range of formats and for objects from the repository to be transformed before dissemination into a format which may be very different from the stored form.

The implementation of the repository uses modern concepts of distributed objects. For interaction with other systems, it uses [CORBA](#) (the Common Object Request Broker Architecture) implemented through [Xerox's Inter-Language Unification \(ILU\) system](#). Internally, the prototype repository uses the [SHORE](#) object-oriented database system, developed at the University of Wisconsin.

IBM repository prototype

The IBM repository prototype is in development for the Federal Theatre Project collection [\[3\]](#). IBM's Research Division has donated software and staff time to work with the Library to build this repository from existing, established IBM products. The digital objects will be managed by [VisualInfo](#), a client/server application for document management that supports images, audio, and video,

as well as more traditional text-based documents. The VisualInfo Library Server is an application built in DB2 (a relational database management system) linked to the VisualInfo Object Server. One Library Server can control access to several distributed Object Servers; the Library Server manages relationships between objects and can hold administrative information. Since the Library Server is built on relational database technology, it can be configured in many ways, and customized interfaces can be built. The VisualInfo Library Server currently supports SQL as a protocol for remote queries, but IBM has expressed an interest in providing Z39.50 support.

IBM is currently developing a program that will load the images already scanned into the VisualInfo repository. This program will integrate data from three sources: database records, created at scan time, that contain logical names and values for certain physical metadata fields; information recorded automatically in the TIFF header of image files; and the image data files.

The development of these prototype repositories is at an early stage, and design details will probably change in the light of practical experience, and as the strengths of each approach become clear. The prototype stage is important in testing the Library's model which separates the storage of primary descriptive information (bibliographic records and finding aids, which may describe intellectual items of which only one manifestation is digital) from the digital representation of the items described. The Library's model allows for multiple repositories, but all repositories must coordinate with independent external indexes and support standard external access protocols. One important aspect of tests is to explore how well the different approaches scale, as the number of objects grows and additional formats must be supported.

Even for historical materials, access restrictions must be considered

The repositories will also support experiments in the handling of access restrictions. Restrictions may be necessary for some items not only to comply with copyright law, but also because of conditions set by donors and the general legal concerns of privacy and publicity. Information has been gathered on the status of many items. Some items have not been digitized or have been withheld from distribution because of the need to restrict access or re-use. Some digitized items may be accessible in the Library's reading rooms but not over the Internet. For example, some photographic collections to which a donor retains rights may be available only through the Digital One-Box in the Prints and Photographs reading room until those rights expire. The NDLP has recently hired a lawyer with intellectual property experience who, working with the U.S. Copyright Office and the Library's Office of the General Counsel, will help establish procedures, clarify the legal position for items and collections not clearly in the public domain, and secure explicit permissions where possible.

What does the future hold?

The Library of Congress is proceeding on parallel paths: working to provide access to resources today, using today's technology, in line with its mission to serve Congress and the American public; and participating in cooperative longer-term efforts to develop a distributed architecture for digital libraries and to build archives of cultural resources. Through the lessons learned as it presses ahead on the first path, the Library hopes to play a constructive role as national and international consortia move along the parallel path. The Library's efforts are based on a flexible, distributed, modular architecture built using open standards, with an emphasis on the widest possible access. Where standards do not currently exist, the Library will contribute its experience as input to the development of such standards, in the Internet and World Wide Web community, the telecommunications community, and the library and archive communities. The Library expects to bring its practices in line with relevant standards as they emerge and are implemented.

Among the consortial initiatives in which the Library of Congress is participating are the National Digital Library Federation (NDLF), the Museum Education Site-Licensing project (MESL), and the G-7 program on the Global Information Society.

[National Digital Library Federation](#)

This federation, established in May 1995, is a consortium of fifteen of the country's largest research libraries and archives. The federation is administered through the [Commission on Preservation and Access](#). A long-term goal is the implementation of a distributed digital library accessible across the global Internet. The library will be formed by the conversion to digital form of resources held by libraries and archives and by the incorporation of materials published in electronic form. Issues under current discussion include the development of a phased strategy for building a technical framework for interoperability and of a taxonomy of resources and rights that must be supported by this framework.

[Museum Education Site-Licensing \(MESL\) project](#)

This two-year project is sponsored by the Imaging Initiative of the Getty Art History Information Program. The goal is to develop a structure for providing images for educational use under terms and conditions that protect the rights of owners while encouraging access. Seven collecting institutions, including the Library of Congress, are providing images and associated descriptions to a common collection, mounted for distribution (currently from the University of Michigan) to seven educational institutions. The Library has already contributed the digital reproductions of [photographs by Carl Van Vechten](#) and will be contributing more images in the next few months, for a total of about 2,000. The Library hopes that the licensing model and distribution structure developed by the project may

allow the dissemination of some historical materials for which copyright restrictions currently prevent uncontrolled access.

G-7 program on the Global Information Society

This program of the G-7 nations (the United States, Canada, Britain, France, Germany, Italy, and Japan) has initiated [eleven pilot projects](#), in areas such as healthcare, emergency management, government information, and electronic libraries. The eleven project themes represent areas where international cooperation will be an asset in demonstrating the potential and stimulating the deployment of global access to information. The Library of Congress represents the United States on the Electronic Libraries Project, which proposes to build a prototype global library ([Bibliotheca Universalis](#)). The Library has recently contributed digitized materials for an exhibit on the theme of Africa to be assembled for introduction at an upcoming meeting.

Other cooperative ventures focus on providing wider and more effective access to the Library's historical collections for the K-12 educational community. Recently, the Center for Children and Technology (CCT) at the Education Development Center in New York City has developed a set of criteria and a related survey form for assessing the educational value of collections. The criteria are grouped in five broad areas:

- U.S. history curriculum content
- Historical thinking skills
- Social studies frameworks (curriculum areas)
- Language arts skills and content
- Teacher usefulness

Each collection has been assessed by a teacher or educational specialist selected by CCT as having the appropriate content expertise and the skills and equipment to access the materials online. Materials from the report detailing the process for selecting the criteria and from the assessments will be integrated into the [NDLP Learning Page](#) over the coming months.

Another activity reaching out to the K-12 educational community is a symposium (in early June, 1996) for educational publishers, co-sponsored with the Association of American Publishers. The NDLP is exploring the idea of forming partnerships with publishers who wish to assemble and deliver material selected from the digitized historical collections in a format and context accessible to elementary and secondary schools.

The Library plans to extend its active support for the digitization of historical materials beyond its own collections. The first project of this type was announced on April 18, 1996. The Library and Ameritech will establish a [competitive grant program to fund digitization of Americana collections at other institutions](#) for

incorporation into the NDLP program. Ameritech will contribute \$2 million over three years to establish this competition.

In parallel with external ventures and collaborations, the Library of Congress continues to select and digitize materials from its own collections. For a preview of collections that should be available soon, see a list of [Future American Memory Titles](#). As digitization and production proceed for these and the other collections already selected for American Memory, the Library of Congress will address, through a combination of its own efforts and the creativity of other institutions and communities, the challenges presented in this article. The Library believes that its experiences will contribute to the realization of the vision shared by so many for a global architecture for digital resources that can be used by the world's libraries and archives to provide widespread access to the treasures they hold for posterity.

□

□

hdl://cnri.dlib/may96-c.arms