Illinois State University
ISU ReD: Research and eData

Faculty and Staff Publications – Milner Library | Milner Library

7-1-2006

# Accessing Digital Libraries: A Study of ARL Members' Digital Projects

Chad Kahl
*Illinois State University*, cmkahl@ilstu.edu

Sarah C. Williams
*University of Illinois at Urbana-Champaign*, scwillms@illinois.edu

Follow this and additional works at: http://ir.library.illinoisstate.edu/fpml

Part of the Library and Information Science Commons

Recommended Citation
Kahl, Chad and Williams, Sarah C., "Accessing Digital Libraries: A Study of ARL Members' Digital Projects" (2006). *Faculty and Staff Publications – Milner Library.* Paper 10.
http://ir.library.illinoisstate.edu/fpml/10

This Article is brought to you for free and open access by the Milner Library at ISU ReD: Research and eData. It has been accepted for inclusion in Faculty and Staff Publications – Milner Library by an authorized administrator of ISU ReD: Research and eData. For more information, please contact ISUReD@ilstu.edu.

# Accessing Digital Libraries: A Study of ARL Members' Digital Projects

Chad M. Kahl, Illinois State University, Milner Library, cmkahl@ilstu.edu

Sarah C. Williams, Illinois State University, Milner Library, ssteini@ilstu.edu

## Abstract

To ensure efficient access to and integrated searching capabilities for their institution's new digital library projects, the authors studied web sites of the Association of Research Libraries' (ARL) 111 academic, English-language libraries. Data was gathered on 1,117 digital projects, noting library web site and project access, metadata and project types.

## Introduction

Before fully implementing a digitization program and investing financial and personnel resources, Illinois State University's Milner Library began discussion on managing digital resources in September 2003. The Digital Resources Management (DReM) task force was created to lead this discussion, and to explore opportunities to lead the campus in coordinating storage, cataloging and retrieval of digital items, while also encouraging the development of new digitization projects. Sub-groups of DReM investigated issues related to metadata, hardware and software, copyright, selection, and access.

Both authors were members of the "DReM Team" and its access sub-group. The main goal of the access sub-group was ensuring efficient access to and integrated searching capabilities for new digital library projects. To meet this goal, Association of Research Libraries (ARL) digital library projects were reviewed and data were gathered in order to benefit from the experiences of other libraries that had already developed digital library programs and projects.

An initial exploration involved twelve Big 10 and University of California system libraries. Due to the useful information found in this initial exploration, the review was expanded to include all academic, English-language ARL libraries. Data were gathered on 1,117 digital library projects, noting library web site and project access, metadata and project types.

This paper presents the study results and notes ways in which Milner Library has used the information to guide its digitization efforts.

## Review of the Literature

A review of library science and information technology literature did not find a publication that gathered data as extensively as this study. No broad overviews regarding digital library project access and searching were found, although publications describing particular digital library projects were available. Christie Stephenson's article covered findings in areas similar to this study—descriptive metadata, database design, interface design and tools for use—but focused solely on the Museum Educational Site Licensing Project.[1] As this study includes a wide variety of projects, it is unique in the literature.

### Research Questions

Before starting this study, the authors had a concept of an ideal digital library project in terms of accessibility. First is an obvious link to the digital library project(s) on the library's home page with intuitive placement and easily understandable language. Second, the project would also be accessible through the library catalog and, if available, through a federated search engine. Third, the ideal project would be both browsable and searchable to provide for different access needs. Fourth, detailed, standard-compliant metadata would be available to aid the discovery process. Lastly, the project(s) would involve a variety of formats, such as image, text, audio, and video.

This study assesses the ideal vision through data gathered on ARL digital library projects. It also answers the following questions:

- How are digital library projects accessed from library web sites?
- What tools and browsing/searching features are available to access the content of digital library projects?
- How frequently is metadata provided in digital library projects?
- What types of digital library projects have been completed?

## Methodology

### *Study Population*

ARL libraries were studied because larger research universities have been significant players in seeking funding for and creating digital library projects. ARL consists of 123 academic and non-academic libraries in Canada and the United States. For this study, the scope was limited to 111 academic, English-language institutions. This limitation was made for two reasons. First, academic institutions have a mission and constituents similar to Milner Library. Second, French-language institutions were excluded to avoid a potentially inaccurate review of their digital library projects, because neither author reads French.

### *Data Collection*

Data gathering was designed to most closely approximate the experiences of a typical patron who uses a library's home page to search for information and resources. After following a link from ARL's list of member libraries to a library's home page, the library's web site was explored to find digital library projects. If one or more projects were found, each was explored and the library's catalog was searched. A library was declared to have no digital library projects if none were found on the library's home page,

resources, special collections or archive pages, site map or from searching the library and university web sites.

Data gathering was done in five phases. The initial review for the DReM report focused only on twelve Big 10 and University of California system libraries. Based on that experience, the data points were formalized; the study was expanded to include an even greater number of ARL libraries; and a MS Access database was created to store the data and allow for future analysis. The second phase of data gathering focused on identifying the number of completed digital library projects at each academic, English-language ARL institution. A total of 1,286 projects were identified at 111 libraries. In-depth analysis of digital library projects in the last three phases led to the elimination of 169 projects, based on the data qualifications that are discussed in detail later in this paper. The final total of digital library projects was 1,117.

(insert table 1)

Data gathering was divided equally between the authors. In order to ensure systematic data gathering, each author reviewed the other's work on a number of occasions. After the second round of data gathering, each author confirmed the number of projects completed by the other's group of ARL libraries. Special attention was paid to institutions that did not appear to have digital library projects. After the last round of data gathering, the data were confirmed for all projects that were marked as including both text and image. Fourteen projects that had been placed in an "other" category were also removed or re-categorized.

Data were gathered at two different levels—institution and project. Two institutional data points were analyzed. First, the number of digital library projects for each institution was counted. In cases where multiple ARL libraries worked on a project, the university hosting the site was credited. Next, the level

of access from the library web site was evaluated. Two levels of access were noted. First level was home page access. Second level was access from a secondary-level page, home page rollover or pulldown menu. If a home page link was available to even one digital library project, the institution was credited with home page access.

For each individual project, nine types of data were gathered. Data points noted whether the individual project was capable of being browsed and/or searched. In terms of access through a library catalog, a library was credited if the catalog contained item-, sub-collection- or collection-level bibliographic records. The capability to search multiple digital library projects was investigated. Other data points noted if access to a digital library project was restricted and whether the restriction was explained. In addition to denial of access, a project was considered "restricted" if publicly accessible content was limited in some way, such as images available only as thumbprints or with reduced resolution. The availability of metadata in the digital library project was recorded, and if available, the ability to search metadata was noted. No distinction was made for the level of metadata. A picture caption was counted the same as a full Dublin Core set of metadata.

The most detailed project-level data point was the type of digital library project. One or more of the following types were noted: atlas or map, audio, video, exhibit, finding aid, image (e.g. photographs, postcards, slides, but not images from scanned pages of documents) and text (e.g. scanned documents, e-journals, e-books). In addition, any project of two or more types was highlighted.

Additional data were collected, but not analyzed. At the institutional level, a field was available for the terminology used by the library or university to identify digital library projects; examples of terminology include digital library, electronic archive, digital resources, online collections, and digitization projects. At the project level, both on- and off-campus partnerships were recorded. URL and notes fields were provided for both project- and institutional-level data entry.

The initial plan was designed to note whether a digital library project web site had undergone usability testing. It quickly became evident after the first two rounds of data gathering that little information was available on usability testing. Anecdotally, references were made to usability studies or testing in less than a handful of institutions.

The authors also planned on determining whether digital library projects were an element in a federated search. As with usability testing, concurrent searching of digital library projects with traditional resources in the library catalog (e.g. books, government documents, audiovisual materials) and article and reference databases occurred too rarely in the study population to include as a data point.

During data collection, considerations were made for adding the following data points: the size of the collection, the level of metadata, and type of digital library software. Each of these was eventually abandoned due to a variety of reasons. The size of the collection was often impossible to determine, because many of the digital library software programs do not gather all the project items into one browsable view. More significantly, the amount of time to count the items would have prohibitive. The level of metadata was not gathered because of time constraints and because neither author has a cataloging background. Tracking the types and numbers of digital library software was also considered, but it was often too difficult to ascertain, since the software was not always clearly identified and was often seamlessly integrated into an overall digital library interface and therefore invisible to end users.

There were some limitations on the ability to gather data. Multiple efforts were made to access digital library projects that initially registered broken links. In cases of restricted access digital library projects, many of the project-level attributes (e.g. browsing and/or search capabilities) were unable to be ascertained. Restricted access also artificially lowered the number of project types.

*Data Qualifications*

Not surprisingly, given the wide variety of institutions and digital library projects, it was necessary to establish qualifications for data gathering. The following types of resources were not evaluated: commercial products; digital library projects only available from or housed in branch libraries, consortial home pages or digital library project indexing pages; e-reserves; database and web site indexing databases; individual e-journals, if main journal indexing page was available; beta, experimental, pilot and "in progress" sites; statistical datasets (e.g. GIS, ICPSR); mirror sites; text-only exhibits and finding aids; bibliographic databases without additional content (e.g. images, audiovisual files); and informational web sites, such as a text-only history of the library.

These qualifications were made for a variety of reasons. Since the purpose of this study was to learn from projects developed by ARL libraries, commercial products were excluded. Due to time constraints, branch library and consortial projects were not explored. Beta, experimental, pilot and "in progress" sites were not evaluated because only completed projects were of interest. Other types of resources, such as statistical datasets, text-only exhibits, and bibliographic databases without additional content, were excluded because they were not relevant to the initial project proposals at Milner Library.

## Results

*Prevalence of Digital Libraries*

Of the 111 academic, English-language ARL libraries, 89 (80.2%) had digital library projects. A total of 1,117 digital library projects were identified at the 89 libraries—an average of 12.6 digital library projects. The median number of digital library projects was eight. The large difference between the average and median occurred because approximately half of the digital library projects were created by

only fourteen institutions. The institution with the greatest number of digital library projects was the University of Michigan with 112. Factoring in the 22 academic, English-language ARL libraries without digital library projects, the average number of projects dropped to 10.1 and the median number of projects was six.

*Attributes of Digital Libraries*

Library Web Site Access

Fewer than half of digital library projects (43.8%) were accessible from the library's home page. For example, the home page of the University of Utah (http://www.lib.utah.edu/) has a *Digital Collections* link under *Research Tools*. This percentage would have been lower if institutions had not been credited with home page access for linking even one project from the home page. More than one out of every five digital library projects (20.2%) were not accessible from the home page or from a secondary-level page, home page rollover or pulldown menu.

(insert table 2)

Project Access

The predominant form of access provided was browsing (91.4%). The option to search was available for slightly more than half (51.5%); this included any type of searching within the project, such as keyword, subject, title and author. Fewer than half (45.7%) provided both search and browse options. For instance, the interface of University of Toronto's *The Discovery and Early Development of Insulin* (http://digital.library.utoronto.ca/insulin/) allows the user to search and browse the collection at any time from a right-side menu.

Some projects were searchable through a library catalog or with other digital library projects. Only one-third (33.3%) of the digital library projects were available in the library catalog; a project was credited with library catalog access for collection, sub-collection or individual item bibliographic records. Slightly more than one-quarter of the digital library projects (25.5%) could be searched simultaneously with other digital library projects from that institution, using digital management software. Notably, the University of Minnesota's Twin Cities' *Images* digital library project (http://digital.lib.umn.edu/advancedsearch.phtml) allows concurrent searching of any combination of its digital collections.

Very few digital library projects had restricted-access (2.9%). The low percentage was likely due to libraries self-selecting projects that do not entail additional barriers, such as copyright concerns. University of Virginia's *Early American Fiction Collection (1789-1875)* (http://etext.lib.virginia.edu/eaf/) offered different levels of access to texts for authorized users versus public users. In addition, the browsing indexing page for the authorized users noted the restricted collections. As a result, public users were given two indications that they could not access particular texts.

(insert table 3)

Metadata

Over one-quarter of projects (27.1%), typically exhibits, did not contain metadata. Metadata ranged from as little as file size or title to forty-plus fields. Less than half (43.6%) allowed metadata to be searched. Washington State University's *Early Washington Maps: A Digital Collection* (http://www.wsulibs.wsu.edu/holland/masc/xmaps.html) is a project that provided very detailed and

extensive metadata. As a result, the digital management software used nearly every word in the title, creator, description, subject and location as links to related resources.

(insert table 4)

Project Types

Digital library projects overwhelmingly consisted of either image or text content—60.8% for image and 57.5% for text. The remaining types of content occurred much less frequently—atlas or maps (6.5%), finding aids (4.9%), audio (4.3%) and video (3.6%).

Nearly one-third of the projects (32.2%) offered content of two-or-more types. Text and image digital library projects were the most common (23.5%). Audio and video digital library projects occurred very rarely (0.9%).

It is worth noting that exhibits were slightly different, because they are a format, whereas the other digital library type designations are based on content. Every project identified as an exhibit also had at least one distinct content type. Image content was most common in exhibits—over four-fifths (83.2%) of the exhibits included images. In general, exhibits were popular accounting for one-fifth of the digital library projects (19.7%).

(insert table 5)

## Conclusion

*Application at Milner Library*

Based, in part, on the findings from the authors' research, Milner Library made the following decisions regarding the presentation of its digital library projects.

First, a link will be provided to the digital library projects from the home page. The center of the home page currently has four rectangular sections, stacked vertically, titled *find articles on your topic* (linked to the article and reference database page); *find books and other materials* (linked to the online catalog); *find a known article* (linked to our link resolver); and *find course reserves* (linked to the course reserves module of the online catalog). When the final versions of the digital library collections are completed, access will be provided in a fifth rectangular section titled *find digital collections*. The term "digital collections" was selected from a list of ten frequently utilized terms for digital library projects noted in the authors' work: digitized collections, digital collections, digital resources, digitization initiatives, digital library initiatives, digital library activities, digital library projects, digital library, online collections and digital archives. The term will undergo usability testing to ensure it is a readily understood term by patrons.

Second, Milner Library belongs to a state consortium that recently licensed a federated search engine. Initially, Milner Library's digital library projects will not be included in the federated search engine, but based on this research, the addition of these projects will certainly be investigated as a means to increase access to their content.[2]

Third, the draft web page for the digital library collections (http://tempest.lib.ilstu.edu/index.php) provides both searching and browsing options for Milner's digital library projects. The authors' recommendation, noted in the DReM final report, influenced the choice to license CONTENTdm software, in part, due to it providing both access options.[3]

Fourth, standards-based metadata is provided for all items in the digital library projects. The metadata standards were chosen independently of the authors' research.[4]

*Future Directions of Study*

While gathering the data, the authors remarked again and again, "that would be interesting to study…" To support further research, our digital library database is available.[5] The authors simply ask you to reference the resource.

Data points that were abandoned during the course of work could be completed by others—the size of the collection, the level of metadata and type of digital library software.

Much more detailed research could use the digital library project database data. Differences between public and private ARL libraries or institutions in the same state or consortium could be examined. One may want to focus on exploring digital library projects of a particular type in greater details. Perhaps a review of every digital library project that combines audio and video offerings could be done. Exploration could be made of what effects, if any, the level of online catalog records has on research. Will individual item bibliographic records lead to greater use than collection- or sub-collection-level records?

Use could also be made of data that was gathered but not analyzed, such as the on- and off-campus partnerships for projects or the digital library terminology for institutions. Perhaps the data could be expanded to include non-English language and/or non-academic ARL libraries.

This data may also be used to identify digital library projects for more detailed, qualitative analysis. For instance, at the ALA poster session presenting findings from a subset of this data, two people asked about specific issues related to policy and decision-making. Even though that was not the focus of this study, others could use this data to perform that type of research.

Since many digital library projects are consortial efforts, examination of these digital library projects would offer a wealth of opportunities. Furthermore, the consortial efforts often involve a wide variety of institutions, such as public libraries, historical societies, museums, archives, etc. Examples of large statewide consortia can be found in California—*California Digital Library* (http://www.cdlib.org/); Florida—*Publication of Archival, Library & Museum Materials* (http://palmm.fcla.edu/index.html); Georgia—*Digital Library of Georgia* (http://dlg.galileo.usg.edu/); Kentucky—*Kentuckiana Digital Library* (http://kdl.kyvl.org/); Louisiana—*LOUISiana Digital Library* (http://louisdl.louislibraries.org/); and Ohio—*Ohio Memory* (http://www.ohiomemory.org/).

**Notes**

1. Christie Stephenson, "Recent Developments in Cultural Heritage Image Databases: Directions for User-Centered Design," *Library Trends* 48, no. 2 (Fall 1999): 410-37.

2. Andy Taylor, telephone conversation, December 1, 2005, and Beth Schobernd, conversation, December 5, 2005.

3. Morag Boyd, telephone conversation, November 18, 2005.

4. Morag Boyd, conversation, November 18, 2005; Morag Boyd, email, November 28, 2005; and Virginia Kerr, email, November 28, 2005.

5. Please contact either author for the data. In addition to this data, there are a number of useful indexing sites to publicly accessible digital libraries: Public Access Collections (Washington, DC: Council on Library and Information Resources, Digital Library Federation), http://www.hti.umich.edu/cgi/b/bib/bib-idx?c=dlfcoll; ARL Digital Initiatives Database (Washington, DC: Association of College and Research Libraries), http://www.arl.org/did/; Inventory of Canadian Digital Initiatives (Ottawa, ON, Library and Archives Canada), http://www.collectionscanada.ca/initiatives/.

**Table 1**

**Data Collection Phases**

| Time Frame | Task | Number of Institutions | Number of Projects |
|---|---|---|---|
| November-December 2003 | Initial review to formalize data points. | 12 | 14 |
| January 2004 | Initial determination of the number of digital library projects at qualifying ARL institutions | 111 | 1,286 |
| February-March 2004 | Gathered data on projects at institutions with 15 or fewer digital library projects | 58 | 417 |
| April-May 2004 | Gathered data on projects at institutions with 16 to 25 digital library projects | 19 | 193 |
| July-August 2004 | Gathered data on projects at institutions with 26 or more digital library projects | 12 | 507 |

**Table 2**

**Library Web Site Access**

| Accessibility | Number (n = 89) | Percentage |
|---|---|---|
| Home Page | 39 | 43.8% |
| Secondary-Level Page | 32 | 36.0% |
| Home Page or Secondary-Level Page | 71 | 79.8% |

**Table 3**

**Project Access**

| | **Number (n = 1117)** | **Percentage** |
|---|---|---|
| Browsable | 1021 | 91.4% |
| Searchable | 575 | 51.5% |
| Browsable & Searchable | 511 | 45.7% |
| Accessible from Library Catalog | 372 | 33.3% |
| Searchable with Other Digital Library Projects | 285 | 25.5% |
| Restricted Access | 32 | 2.9% |

| **Table 4**<br><br>**Metadata** | | |
|---|---|---|
| | **Number (n = 1117)** | **Percentage** |
| Metadata Included | 814 | 72.9% |
| Metadata Searchable | 487 | 43.6% |

**Table 5**

**Project Types**

| Type | Number (n = 1117) | Percentage |
|---|---|---|
| Image | 679 | 60.8% |
| Text | 642 | 57.5% |
| Image & Text | 263 | 23.5% |
| Exhibit | 220 | 19.7% |
| Atlas/Map | 73 | 6.5% |
| Finding Aid | 55 | 4.9% |
| Audio | 48 | 4.3% |
| Video | 40 | 3.6% |
| Audio & Video | 10 | 0.9% |