2005

An Integrated Framework for Discovering Digital Library Collections

Karen S Calhoun, *University of Pittsburgh - Main Campus*

An Integrated Framework for Discovering Digital Library Collections

KAREN CALHOUN

Library Technical Services, Cornell University, Ithaca NY 14853

E-mail: ksc10@cornell.edu

Abstract: Information seekers are generally on their own to discover and use a research library's growing array of digital collections, and coordination of these collections' development and maintenance is often not optimal. The frequent lack of a conscious design for how collections fit together is of equal concern because it means that research libraries are not making the most of the substantial investments they are making in digital initiatives.

This paper proposes a framework for a research library's digital collections that offers integrated discovery and a set of best practices to underpin collection building, federated access, and sustainability. The framework's purpose is to give information seekers a powerful and easy way to search across existing and future collections and to retrieve integrated sets of results.

The paper and its recommendations are based upon research undertaken by the author and a team of librarians and technologists at Cornell University Library. The team conducted structured interviews of forty-five library staff members involved in digital collection building at Cornell, studied an inventory of the library's more than fifty digital collections, and evaluated seven existing OAI and federated search production or prototype systems.

The author will discuss her team's research and the rationale for their recommendations to: present a cohesive view of the library's digital collections for both browsing and searching at the object level; take a programmatic (rather than project-based) approach to digital collection building; require that all new digital collections conform to library-developed and agreed-upon OAI best practices for data providers; and implement organizational structures to sustain the library's digital collections over the long term.

Key words: Digital collections, federated searching, OAI, digital library best practices

BACKGROUND

Cornell's situation vis-à-vis digital collection dispersion is worse than at most institutions because Cornell was a digitization pioneer and created digital content long before relevant metadata and other standards emerged. At this stage, no new effort should be undertaken without a sense of how it will be merged with other existing collections and where the resources for long-term maintenance will come from. —A Cornell digital projects librarian

Currently, users are on their own to discover and use the Cornell University Library's rich array of digital collections, and coordination of these collections' development and maintenance is not optimal. The Library's current lack of a conscious design for how collections fit together is of equal concern because it means the Library is not making the most of the substantial investments it has made (and will make) in digital initiatives. In early 2004 the author and a team of librarians and technologists began exploring how Cornell University Library might develop an integrated technological and methodological framework to make the Library's fifty-odd digital collections easier to discover and use (Cornell University Library, 2004, 2005a). This paper summarizes what they learned. Libraries like Cornell, which began investing in digital collections building over a decade ago, now experience an acute "embarrassment of riches" on the one hand, and near chaos on the other. Assuming education is less expensive than ignorance, other libraries just starting out with digital initiatives might benefit from the experiences of early adopters like Cornell.

Besides reporting the observations and insights of experienced digital library specialists, this paper also proposes an integrated discovery framework for digital collections and a set of best practices to underpin collection building, federated access, and sustainability. In particular this paper describes:

> • How an integrated framework should look and function, from a user's perspective

• Implementation issues related to introducing and supporting an integrated framework
• Recommended best practices
• A possible system architecture (model) for the integrated framework

The reader is advised that the topic of this paper—an integrated framework for discovering digital collections—is difficult to consider apart from other areas of intense investigation at this time, namely:

1. The requirements for general and specialized portals that organize large masses information from multiple sources in ways that make sense to general readers or specialized communities.

2. Issues of long-term preservation and accessibility of digital assets.

3. Mass digitization projects.

4. Electronic publishing, as it relates to building a discipline-based repository or repositories.

Nevertheless, this paper's scope is limited to discussion of the requirements and implementation issues associated with integrating discovery of a particular type of digital asset, the digital collection.

METHODOLOGY

This paper and its recommendations are based upon:

• Notes from structured interviews of forty-five Cornell University Library staff members
• A literature review
• An evaluation of seven OAI and federated search production or prototype systems
• An inventory of the Library's digital collections

The full report (Cornell University Library, 2004) contains details of the interviews, including the names of those interviewed and a summary of the results; a spreadsheet that summarizes the evaluation of the seven systems; and other related materials.

HIGH-LEVEL REQUIREMENTS FOR THE INTEGRATED FRAMEWORK

Users are particularly concerned with issues such as speed, convenience, minimization of hassle, comfort, straightforwardness in use and presentation, etc. For some constituents, these factors are more important that the quality of the collections themselves.—An interviewee

Key requirements

Asked how an integrated framework for the Cornell University Library's digital collections should look and function, interviewees generally agreed that the framework should offer a Web-based, cohesive view of all collections, with easy navigation, cross-collection searching, and the ability to link out to the digital objects thus discovered. System performance (fast response time), easy searching, navigation, and browsing, and good output options ("shopping cart" functionality plus printing, marking, saving, downloading) were mentioned frequently. Another theme of the interviews was "much more like Google," in the sense of faster search engines and better indexing to assure relevant, accurate, consistent query results.

Speed

Those interviewed are not alone when they speak of their dissatisfaction with the response time of federated searching (also called metasearching or distributed searching) (Frost, 2004). At the time the Cornell University Library's federated search application, Find Articles, was introduced in May 2003, average response time to a query of multiple databases was in the range of 10 to 20 seconds. In his book on usability, Jakob Nielsen (1994), reports "the basic advice regarding response times has been about the same for almost thirty years" and notes that a 10-second response time is about the maximum for retaining the average computer user's focus, and one second is about the

maximum for the user's flow of thought to remain uninterrupted. While Find Articles is an integrated framework for discovery and access to remote licensed electronic resources, rather than the Library's digital repositories, those we interviewed seem to fear that federated searching of the Library's digital collections will offer similar response times.

"More like Google"

Google searches generally give results in ¼ second (Sullivan, 2004). Google's lightening fast results happen for two reasons: very large amounts of computer memory and a large cross-file index cached in memory. A Google user is searching an index, not the millions of Web pages used to create that index. Interviewees frequently mentioned "the Google factor" and suggested that the Library's discovery system for digital collections should look and work as much like Google as possible. This is a goal to strive for. At the same time, it is important to develop realistic expectations for how much like Google a research library's cross-collection search services can be (Luther, 2003).

How much of the Google model is transferable to a cohesive view of a library's digital collections? Dispersed, largely text-based databases and diverse metadata are defining characteristics of the domain of scholarly information. A federated search executes a cross-file query across databases that do not share a common index and whose metadata varies from database to database, making rapid response and consistency in result sets more challenging to achieve.

Searching, navigation, browsing, output options

The team found and evaluated a variety of specifications and planning documents for systems intended to federate access to disparate online collections. In the course of this research the team also became familiar with "The European Library" (TEL) project, whose goal is to provide a Web-based service to greatly improve access to the digital and non-digital material held in Europe's national libraries (The European Library, 2005). As will be described later in this report, the TEL project influenced the team's thinking a good deal. The team developed a set of mandatory and desirable functional requirements based on these sources and the team's own experiences. Using the list of requirements, the team then evaluated the early 2004 versions of seven systems that provide federated access to multiple digital collections: RLG Cultural Materials, the University of Washington Digital Collections, OAIster, a locally-built example from the Library's systems office, the National Science Digital Library (NSDL), the New Zealand Digital Library, and the European Library (TEL).

The systems varied widely but shared some common characteristics. Most allowed for browsing, at least by collection. Most allowed for both a basic and advanced search although basic keyword searching was the most common and prominent way of searching. Limiting and sorting options were available on many with three systems allowing for pre-limits and one allowing for post-limiting. The most successful limiting options involved limiting by format (text, image, video, etc.). The ability to sort results was not widely available with only two systems allowing for very limited sorting of results. Output options were weak in these seven systems. Only one system (University of Washington) allowed for the marking and saving of search results and none had the capability to email records (output options were considered highly desirable in the interviews with Cornell University Library stakeholders). More advanced features such as personalization were also lacking in these systems with only the University of Washington allowing for the storing of favorites.

Usability

A strong theme in the interviews was the importance of involving users in defining system requirements and of conducting usability studies of digital collection interfaces. User needs vary from digital collection to digital collection. The research team is proposing a system that allows a library's digital collections to be searched collectively; however it is assumed that users will

retain the ability to search collections individually in their native interfaces, and that separate interfaces will continue to be built for each digital collection. So usability must be defined on (at least) two levels—the usability of each digital collection's native interface, and the usability of the integrated framework. Brinck (2002) and others writing on usability offer guidance that could be applied to profiling potential users of the integrated framework and understanding their needs, levels of expertise, and so on.

Digital collections content

> *So many of the projects today are a hodgepodge of material with little deliberate thought on the scope and completeness. We should find out what the 'hot' topics are for researchers so we are digitizing material that truly meets their needs.—Interviewee*

While most interviewees focused their remarks on how they want the integrated framework to function, a few commented on requirements related to the content of the collections. The interviewee's comment that begins this section is admittedly strong, but it does reflect the assessment of some of the Library's existing content—largely the result of a mix of one-time funding opportunities—is a somewhat eclectic set of mostly smaller or incomplete collections. Other interviewees urged the research team to begin to "think big" when developing content—for example, design projects not in terms of one museum's nineteenth century botanical illustrations, but all of that century's botanical illustrations. An integrated framework could foster clear direction and purpose for a library's digital collection projects, a role for which library selectors and reference staff are well suited.

The Open Archives Initiative (OAI)

The OAI world is divided into *data providers*, who make their metadata available, and *service providers*, who harvest metadata from data providers. Service providers also build services—most often search and retrieval services—around the harvested metadata. The research reported here suggests that, for the benefit of the communities they serve, libraries should play both roles. As a *service provider*, a library federates access to its own local collections—this is the goal of the integrated framework.   At the same time a library will also want to be able to participate in large-scale initiatives as a *data provider* by exposing metadata about the content of its local collections to larger aggregations. One interesting project along these lines is Aquifer, a project of the Digital Library Federation (Kott, 2005). It is important for libraries to begin proactively facilitating the efficient and cost-effective dissemination of its digital collections content. This will be even more important in the future.

Recommendations

Using data from the interviews and the review existing systems, the team developed some assumptions about mandatory user requirements for the integrated framework:

- The system will be web-based.
- The system will present a cohesive view of as many digital collections as is feasible for both browsing and searching.
- The system will allow the digital collections that are incorporated in the framework to be searched collectively while still allowing them to be searched individually in their native interfaces.
- The system will be browseable at a minimum by collection.
- The simple search will be a Google-type box for free-text searching though an advanced search would be desirable. It's well documented that users underutilize advanced features of search systems. At the same time, some reasonable compromise between Google and a system to please a professional searcher is needed. The notion of allowing users to limit their searches by format (text, image, video, etc.) is also important. Getting the right mix of features depends on understanding the needs of a user of the integrated framework, as mentioned in the

"Usability" section.

• Results will be presented in a clean manner for easy scanning.
• Links to user documentation will be made prominent throughout the system though the system should be easy to use by anyone familiar with the Internet.
• Links for obtaining staff help for technical and reference questions will be made prominent.

The team further recommended that the proposed mandatory requirements be further tested and validated through user profiling and needs assessment. The team's other recommendations related to high-level requirements are:

• **Speed**: System response time should be 10 seconds or less for 90% of searches.
• **Usability:** Consult with a usability expert to develop a profile of the prospective user of the integrated framework; build the system using the user profile as a guide; conduct usability studies.
• **Content:** Evaluate and develop criteria for what collections, existing and prospective, should be included in the integrated framework.
• **OAI:** Require all new collections be at a minimum OAI harvestable.
• **OAI best practices:** Require all new collections be in conformance with library-developed and agreed upon OAI best practices for data providers. Continue participation in the DLF OAI Best Practices group (Digital Library Federation and National Science Foundation, 2005). Evaluate what it would take for the integrated framework to serve as both service provider and data provider for dissemination of content to larger-scale efforts such as Aquifer.

IMPLEMENTATION ISSUES

Collection building

The Cornell staff members who have been involved in digital projects are justifiably proud of the wealth of high-quality digital library content they have built. Their results represent over a decade of effort. Projects have been well managed and successful. At the same time, because most projects were externally funded, the collections that came from them tend to be isolated in terms of how they relate to other collections, who is responsible for them, delivery platforms, and how standards are applied.

When asked what has worked well with the Library's digital initiatives, interviewees most often responded:

• When project leadership/management is good
• When we have established workflows/guidelines (when we have experience)
• When the team is cross-functional
• When communications are good

When asked what digital initiatives challenges have been, interviewees pointed to:
• Not enough technical support (and competing for resources to do projects)
• Bad front ends (interface an afterthought; not designing with user in mind)
• Changing technology
• Weak or no ongoing support (after project is done)
• Lack of skills/competencies/commonly endorsed guidelines/methodology
• Poor communication about what we are doing/have done
• Metadata not available for specialized materials

The interviews reflected a tension between anticipated benefits of coordinated digital collection development and wariness that coordination might stifle the innovations of independent digital collection developers. Several interviewees suggested it would be helpful for collection builders to have common shared guidelines, standards, practices, and access to various kinds of checklists. Some wished for more programmatic funding and less reliance on grants. Interviewees often mentioned the need for a clearinghouse for collection

builders, to seek not only advice but also services like "rent a techie" or "rent a project manager." At the same time, some interviewees were nervous about yielding control to a central office. Interviewees would welcome more involvement in collection building by collection development staff and faculty.

Sharing collections: number of delivery platforms

> *No one system will work adequately for all of [the Library's digital collections]. New innovations will build on old or migrate. There will always be new things that we'll want to incorporate.—Interviewee*

The inventory of the Library's digital collections demonstrates that the collections are delivered using a wide variety of platforms. The larger the number of delivery platforms, the higher the cost of building and maintaining an integrated search and retrieval framework. When asked if the number of delivery systems could be reduced, interviewees generally agreed that it would be impossible to reduce the number to one and that the choices should not be narrowed too far, because different materials and different user communities require different systems. Other oft-made points were that it is technologically too soon to reduce the number of delivery platforms, or that doing so could restrict innovation or reduce flexibility in user-centered design. One interviewee captured the general sentiment by saying "our goal should be standardized best practices, not homogenization."

Sharing collections: federated search or OAI?

Interviewees with more technical expertise were asked whether the integrated framework should be modeled on OAI harvesting and metadata aggregation or on federated searching. The most frequent reply was both; the two are not mutually exclusive.

In spite of its popularity with library users, there is some disillusionment with federated searching as implemented in the Library: not all databases can be available for cross-searching; searches are comparatively slow; searching is less precise and results not

comprehensive; it is difficult to sort search results meaningfully; output options are poor (Calhoun, 2004). The OAI model for federated access, with which Cornell Library interviewees have less direct experience, is perceived as less problematic, scalable, faster, and more standards-based. Yet as the OAI protocol has become more widely adopted, service providers have discovered some major harvesting issues (Tennant, 2004). Implementers of the OAI protocol for metadata harvesting (OAI-PMH) are experiencing problems not dissimilar to the ones experienced by federated search implementation teams. The central problem for both kinds of implementers is the diversity of metadata that the integrated framework must aggregate and make sense of for indexing, search, sorting, and display.

Based on the interview comments, OAI is likely to appeal to a library's collection builders. Designed to be a low-barrier interoperability framework, adopting OAI could be the catalyst needed to federate discovery and access across our diverse collections, provided libraries also adopt emerging OAI best practices. A policy to make all future digital collections OAI-harvestable is likely to add little to the cost of digital projects, making conformance to the policy more likely. At the same time, success with OAI depends on a library's collection builders' making and complying with up front agreements on collection- and item-level descriptions and on mechanisms for extending the core metadata when a collection supplies domain-specific fields. Without this consensus, it will be difficult to build meaningful OAI indexes on the set of OAI-harvested metadata.

Sustaining library digital collections

If the research team had to pick one and only one overriding theme from the stakeholder interviews, it would be *sustainability*, often expressed by interviewees as "moving from project to program." Those we interviewed tend to perceive long-term stewardship of the Library's digital collections as weak. They feel more organized support and more resources are needed, along with a more unified approach to decision-making, clearer institutional policy and direction, and better communications. Several

advocated the assignment of high-level project sponsors or advocates and/or a Digital Projects Working Group, where project teams could interact and share tools, crosswalks, ideas, define new policies and practices, etc.

The team recommended that a library's requirements for sustaining access to the Library's digital collections be assessed in light of its strategy for archiving and preservation (for an instructive example, see Kenney, Rieger, and Entlich, 2001). Further, the team recommended support for the "access entity" of the OAIS-based digital archive now under investigation at Cornell.

Institutional learning and documentation

Some of the interviewees' suggestions fell into the category of tools to enhance organizational learning. They include a variety of registries to facilitate communication, capture what project teams learn for the benefit of future project teams, streamline new project startups, and reduce duplication of effort; usage statistics and reports to capture information about user behavior; and a variety of forums to bring digital collections staff together and help them learn from one another.

BEST PRACTICES

A library digital collections program

The concept map (fig. 1) lays out what could become a library digital collections program. The program would consist of an ongoing cycle of building, sharing, and sustaining both individual digital collections and also the integrated discovery framework. A set of best practices for each stage of the cycle appears within each quadrant of the square. For further explanation of the best practices, see Appendix E of the full report (Cornell University Library, 2004).

Besides building, sharing, and sustaining digital collections, coordinated planning is essential and draws extensively on dynamic, well-documented institutional learning. From this foundation of institutional learning, a library can build, share and sustain collections in an informed and coordinated manner, relying heavily on institutional experience, as well as appropriate faculty and user input. In support of data sharing, well-defined policies regarding collection presentation, delivery platforms, reusable metadata (Kurth, Ruddy and Rupp, 2004), and OAI best practices are crucial. Further, project sponsors and assigned collection maintenance staff, intelligent use of administrative and preservation metadata, and a well-documented preservation policy ensure the sustainability of the collections. Compiling, documenting, and communicating information on new and existing collections not only helps to sustain these collections, but also stokes the institutional learning component of the process, thus completing the cycle.

A key to the success of the proposed library digital collections program is adequate funding and manpower. Every stage of the life cycle relies on them. Another critical success factor is an explicit, agreed-upon collection development policy to guide what digital collections are built and how they relate to each other and to national and international digital library initiatives.

Conceptual diagram of the integrated framework

Achieving consensus on library best practices and organizational structures for creating, sharing, sustaining, and documenting digital collections is critical to the success of an integrated discovery framework. Drawing on the structure and principles presented in "A Framework of Guidance for Building Good Digital Collections" (NISO Framework Advisory Group, 2004), the research team identified the most frequent and compelling responses to interview questions and compiled a set of best practices and standards to address the needs of a library's digital collections community. A conceptual diagram of these findings follows (fig. 1). The diagram illustrates two aspects of a proposed library digital collections program. The foursquare box on the left represents the organizational life cycle of building, sharing, sustaining, and planning for new digital collections and the integrated discovery framework. The diagram on the right illustrates

how "sharing" would manifest itself inside and outside Cornell.

In drawing the diagram, the author assumed that the Library wants to include all of its existing digital collections for cross-collection searching. These fifty-odd collections (depicted in the figure as ovals) are a hybrid of OAI-harvestable resources, non-OAI-harvestable resources, and Z39.50 targets with their own separate interfaces, served through a variety of delivery platforms. The numbers in parentheses are the number of existing collections in each category, e.g., 11 Luna Insight collections, 11 collections delivered with Web software, 10 with DPubs, etc. The Web delivery systems are particularly diverse. It is further assumed that the current state of affairs will continue except that the Library may eventually pare down the number of delivery platforms and require all new collections to be, at a minimum, OAI-harvestable in conformance to an agreed upon set of OAI best practices. The legend "Prospects for Integration" at the bottom of the diagram indicates the anticipated ease or difficulty of federating access to the types of existing collections at the time the diagram was made.

Service Provider and Data Provider

The figure depicts one other important concept. The research team recommended that the integrated discovery framework act as a service provider, aggregating metadata from a library's digital collections and providing a single search and retrieval service for anyone interested in the library's digital collection holdings. At the same time, the integrated framework can act as a data provider to a **d**istributed **o**pen **d**igital **l**ibrary (DODL), exposing a library's digital collections metadata for harvesting by other service providers. Participating in cross-institutional partnerships will allow a library to increase its searchable digital collection universe to provide even greater access to the scholarly community—locally, nationally, and internationally.

PROPOSED SYSTEM ARCHITECTURE

Benefits

The team proposed a system architecture for the integrated framework that was influenced by the TEL project and prototype (The European Library, 2005). What follows is an attempt to articulate why modeling the integrated framework on the TEL approach would be beneficial. The TEL approach is novel in that it combines a distributed search model with the OAI model for federating access (Van Veen and Oldroyd, 2004). The TEL approach features an implementation of SRU--search and retrieval via URLs (ZING, 2005). Among the benefits of the proposed approach are:

- Speed. Due to its innovative use of modern browser capabilities, the TEL solution has the promise of mitigating some of the obstacles to rapid response time in a federated search system.
- A workable starting point and some quick wins. Cornell already has some of the required tools to implement the proposed architecture or can get them from TEL or other sources.
- Incrementally moving the Library forward from its existing "islands of information" toward a more open architecture that follows best practices and standards as they emerge.
- Low barriers to contributing to the integrated framework, now and over time. Existing digital collections could be integrated into the framework without having to migrate or transform them in significant ways. New digital collections would be easier to integrate, assuming collection builders adhere to some basic shared best practices.
- Eliminates the need for a central portal based on a single delivery platform.
- Encourages collection builders to provide structured metadata and to be OAI compliant, but also allows us to integrate valuable non-OAI compliant resources into the framework, now and in the future.
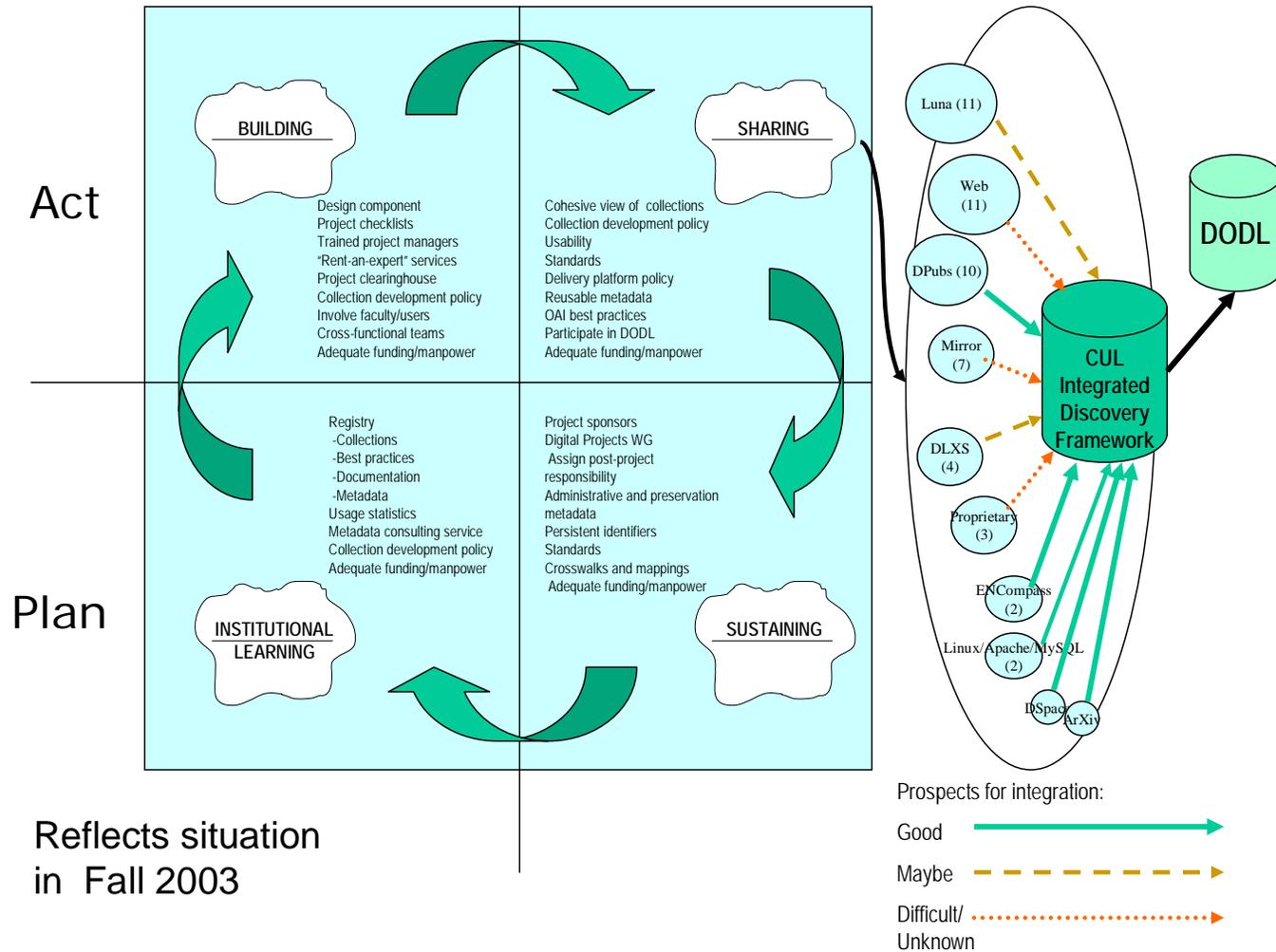
BUILDING

Design component
Project checklists
Trained project managers
"Rent-an-expert" services
Project clearinghouse
Collection development policy
Involve faculty/users
Cross-functional teams
Adequate funding/manpower

SHARING

Cohesive view of collections
Collection development policy
Usability
Standards
Delivery platform policy
Reusable metadata
OAI best practices
Participate in DODL
Adequate funding/manpower

Act

Plan

Registry
 -Collections
 -Best practices
 -Documentation
 -Metadata
Usage statistics
Metadata consulting service
Collection development policy
Adequate funding/manpower

Project sponsors
Digital Projects WG
 Assign post-project
responsibility
Administrative and preservation
metadata
Persistent identifiers
Standards
Crosswalks and mappings
 Adequate funding/manpower

INSTITUTIONAL
LEARNING

SUSTAINING

Reflects situation
in Fall 2003

Luna (11)
Web (11)
DPubs (10)
Mirror (7)
DLXS (4)
Proprietary (3)
ENCompass (2)
Linux/Apache/MySQL (2)
DSpace
ArXiv

CUL
Integrated
Discovery
Framework

DODL

Prospects for integration:

Good

Maybe

Difficult/
Unknown

Figure 1. Digital Collections Program Best Practices

• Allows for the possibility of including Z39.50 targets (such as library catalogs) in the framework.

Alternatives considered

The proposal to use the TEL model and SRU as starting points is only one possibility for implementing an integrated framework for its digital collections. Other possibilities that the team considered include:

- One central portal using one delivery platform. This approach was seriously considered but eventually rejected because of the cost of migrating collections, purchasing commercially provided XML gateways to other delivery platforms, and so on.
- A single repository of OAI-harvested metadata. This approach would have been attractive to many of those the team interviewed, but in the end a single OAI repository was rejected because less than half of Cornell's existing digital collections are now or will soon be OAI-harvestable. Metadata migration, conversion, and maintenance costs of this option could be high. While some felt searching a single OAI repository could be faster and allow for better ranking of search results, our test searching of the implemented OAI repositories such as OAIster did not confirm this view.
- Do nothing. The team also considered the possibility of recommending that the Library do nothing, that is, not build an integrated framework for its digital collections. One reviewer of an early draft of the team report wondered if an integrated discovery framework would add sufficient value, considering that despite the impressive volume of material that the Library has digitized, most of the existing collections are fragments of larger corpuses or otherwise narrow in scope.

Another reviewer was concerned that if the Library cannot build a framework that successfully competes with Google response time and other popular features, it should not build anything at all. This perspective is related to the point made earlier in this report that a good fast interface is often more important to users than the content of what is being searched.

While in the end the notion of doing nothing was rejected, it is crucial that (1) future digitization efforts be guided by proactive mechanisms for identifying worthy collections based on demand, usage trends, current research interests, etc. and (2) the integrated framework's performance be optimized to meet the needs of the particular set of users to whom the Library offers this service.

Conclusion

Libraries as a whole face an enormous challenge positioning themselves in the information market, which is now so dominated by Google and the other Internet search engines. However, it is well documented in the marketing literature that any organization can differentiate what it offers from what competitors offer. The solution is to recognize that information users have different needs and thus can be attracted to different offers. From a marketing perspective, a digital collections discovery framework will succeed to the extent that a library can:
• identify the set of information users to whom the integrated framework can offer important and highly valued benefits
• distinguish what the framework offers from what is offered to this group of users by others in the information market
• provide services to this set of users that are superior to other ways they have to obtain the information
• make the service visible and successfully communicate the benefits
• preempt the content that others can offer
• be perceived as affordable and easy to use
• adequately fund and staff the service

Update on progress toward an integrated framework

In the spring of 2005, the Cornell University Library initiated a project to implement an integrated framework for an initial set of its digital collections (Cornell University

Library, 2005b). Besides exploring an SRU-based approach to federated searching, the implementation group is working on a set of best practices to underpin CUL digital collection building, federated access, and sustainability. The implementers have selected the IMLS DCC Collection Description Schema (Institute of Museum and Library Services, Digital Collections and Content, 2003) for the Cornell University Library digital collections registry. In addition, among other accomplishments to date, the implementation team has developed a proposal to streamline the selection and management of delivery platforms for digital collections and taken the first steps toward assessing digital collection user preferences and behavior.

References

Brinck, Tom, Gergle, Darren and Wood, Scott D. 2002. Usability for the Web. Morgan Kaufmann, San Francisco.

Calhoun, Karen. 2004. Catalogs and Portals: Models and Tools for Discovery and Access. Presented at the "Portals Puzzle" program, ALCTS Catalog and Form and Function Interest Group, June 27, 2004, Orlando, Florida, p. 12-13. Available: http://dspace.library.cornell.edu/handle/1813/1482

Cornell University Library. 2004. An Integrated Framework for Cornell University Library Digital Collections: High-level Requirements and Internal Implementation Issues. Available: http://www.library.cornell.edu/staffweb/integrated_frameworks/Digital_Collections_Framework_Report8Final.pdf

Cornell University Library. 2005a. Digital Collections. Available: http://www.library.cornell.edu/about/digital_collections.html

Cornell University Library. 2005b. CUL Integrated Framework. Available: http://gould2.mannlib.cornell.edu:8080/confluence/display/CULIF/Home

Digital Library Federation and National Science Digital Library. 2005. OAI Best Practices Working Group. Available: http://oai-best.comm.nsdl.org/cgi-bin/wiki.pl?OAI_Best_Practices

Frost, William J. 2004. Do We Want or Need Metasearching? Library Journal 129 (6): 68. Available: http://www.libraryjournal.com/article/CA405394.html?display=searchResults&stt=001&text=metasearching

Institute of Museum and Library Services, Digital Collections and Content. IMLS DCC Collection Description Metadata Schema. 2003. Available: http://imlsdcc.grainger.uiuc.edu/CDschema_overview.htm

Kenney, Anne, Rieger, Oya, Entlich, Richard et al. 2001. Preserving Cornell's Digital Image Collections: Implementing an Archival Strategy. IMLS Project Report. Available: . http://www.library.cornell.edu/imls/IMLS-CULfinalreport2.pdf

Kott, Katherine. 2005. DLF Aquifer: Tapping New Sources for Scholarship. CLIR Issues 45. Available: http://www.clir.org/pubs/issues/issues45.html#dlf

Kurth, Martin, David Ruddy and Nathan Rupp. 2004. Repurposing MARC Metadata: Using Digital Project Experience to Develop a Metadata Management Design. Library Hi Tech 22 (2): 153-165.

Luther, Judy. 2003. Trumping Google? Metasearching's Promise. Library Journal 128 (16): 36-39. Available: http://www.libraryjournal.com/index.asp?layout=article&articleid=CA322627&publication=libraryjournal

NISO Framework Advisory Group. 2004. A Framework of Guidance for Building Good Digital Collections. 2nd edition. National Information Standards Organization, Baltimore, Md. Available: http://www.niso.org/framework/framework2.html

Nielsen, Jakob. 1994. Usability Heuristics. In Usability Engineering. Morgan Kaufmann, San Francisco, p. xxx.

Reference Model for an Open Archival Information System (OAIS). Blue Book. Issue 1. January 2002. p. 4-15. NASA, Washington DC. Available: http://ssdoo.gsfc.nasa.gov/nost/wwwclassic/documents/pdf/CCSDS-650.0-B-1.pdf

Sullivan, Rob. 2004. Why is Google So Fast? Marketing Monitor, January 13. Available:

http://www.searchengineposition.com/info/Articles/GoogleFast.asp

Tennant, Roy. 2004? Bitter Harvest: Problems & Suggested Solutions for OAI-OMH Data & Service Providers. California Digital Library. Available: http://www.cdlib.org/inside/projects/harvesting/bitter_harvest.html

The European Library. 2005. About Us. Available:

http://www.theeuropeanlibrary.org/portal/index.htm

Van Veen, Theo and Oldroyd, Bill. 2004. Search and Retrieval in The European Library. *D-Lib Magazine* 10 (2). Available: http://www.dlib.org/dlib/february04/vanveen/02vanveen.html

ZiNG. Z39.50 International: Next Generation. 2005. Available: http://www.loc.gov./z3950/agency/zing/