

DLF-Aquifer Services Initial Planning Report

Aquifer Services Working Group Initial
Report, Including Information
Gathered to Date, Survey Plan, Use
Cases, and Next Steps

Version 0.7

Last revised June 14, 2005

Table of Contents

Executive Summary	3
Background.....	5
Working Group Charge.....	5
Immediate tasks	5
Services Working Group Members	5
Analysis of Relevant DLF Activities	6
Survey of DLF Aquifer Institutions	6
Methodology	6
Initial Assumptions	6
Initial Responses.....	6
User Studies conducted by DLF Members	8
OAI Usability and K-12 Teachers.....	8
Semantic Clustering.....	8
Digital Resources in Undergraduate Education	9
American West Project User Studies.....	9
Other relevant Studies and resources.....	10
Web Design for Cultural Heritage Materials Presentation	10
Digital Resources in Humanities and Social Sciences	10
Aquifer Use Cases.....	12
End-User Use Cases	12
Use Case #1. Finding Research Materials.....	12
Use Case #2. Browsing Unfamiliar Research Archives	13
Use Case #3. Discovering New Content through Alerting Services	14
Use Case #4. Collecting Items in a Persistent Exportable Basket.....	15
Use Case #5. Searching Full Text	16
Use Case #6. Annotate / Comment.....	16
Use Case #7. Accessing Content Seamlessly through Course Management Systems.....	17
Institutional Use Cases	18
Use Case #8. Metadata Harvesting.....	18
Use Case #9. Metadata Remediation.....	19
Use Case #10. Metadata Enhancement	20
Use Case #11. Taxonomy Assignment.....	21
Use Case #12. Focused Crawling	22
Recommendations and Next Steps	23
Appendix A: Additional Sources Consulted	24
Appendix B: Sample Response to Aquifer Institutional Survey	25

Executive Summary

The Aquifer Services Working Group (SWG) has acted to initially identify, analyze, and articulate use cases for DLF Aquifer services based on an environmental scan. This initial report describes key services that we believe are worthy of attention, and lays out a series of next steps to further refine these service descriptions into functional requirements. Our activities have been guided by the following planning efforts which the group is undertaking:

- The Services Working Group is conducting a survey of DLF institutions to gain an understanding of the perspectives of all our members on the service opportunities and needs to be addressed in the Aquifer project.
- We are carefully reviewing noteworthy studies undertaken by DLF institutions and other agencies of digital library services gaps, opportunities, and potential functions for key stakeholders and constituencies.
- The Services Working Group has used the survey results and studies reviewed to identify a set of digital library services for the Aquifer project in the form of use cases.

We have some early results of the survey to report here, and we aim to finish all survey activities by the end of summer 2005. While these analysis efforts are still at a relatively early point, we have immediately seen many basic areas of consensus that led us to a logical set of services for initial consideration. The following are key points to make about these services:

- The Aquifer project presents an opportunity for DLF institutions to create new services collaboratively that would be difficult or impossible for them to create individually. It is likely that this will require that we think about services in new ways, as coherent systems made up of composite modular functions operated by respective institutions acting in concert to achieve specified goals.
- While Aquifer content may be discovered and accessed through a central interface, we assume that ultimately end-users will discover and use Aquifer content through the interfaces, systems and tools that their local research library provides in ways that seamlessly integrate this content with other content developed, managed or licensed by those institutions. This will require the elaboration of a corresponding set of institutional services and use cases. This is the essential aspect that emerges from the metaphor of an Aquifer: a pooled resource held in common, shared through a hidden layer, and which manifests itself in particular locations in various forms.
- Because of the distributed nature of the Aquifer, its services will be aimed at both end-users and DLF institutions. This shared focus on institutions as well as end-users is important, as there are many enhancement functions for digital objects and metadata which individual DLF libraries would like to be able to accomplish, but which are only probable in a collaborative setting. We have sought to analyze services broadly during this early period of planning in order to identify as many collaborative opportunities as possible which have not been considered previously.

- Our basic approach here is use case analysis, a technique that is commonly used in systems design in both research and industry. We feel that this provides a concrete idea of how the services would work and what needs each would address. While our focus is on Aquifer Phase 1, we have also identified use cases that may pertain to subsequent phases of the project as their sophistication may require significant planning that should start soon.
- Virtually all of the use cases that we describe in this report can and should be implemented at several levels. Variations on most of these use cases could usefully be implemented as either centralized services for all of the Aquifer collections or as localized services in the context of individual institutions.
- Virtually all of these service variations would benefit from collaborative development efforts, but will also need particular champions to assume the responsibility for carrying them forward. Active participation by all Aquifer institutions will therefore be essential to the project's success. There will clearly be pride of place positions that emerge for *all* Aquifer institutions in the implementation of the range of services we have identified. This is not surprising upon reflection given the scale of the development challenges we face in creating the next generation of digital library services.
- What we find emerges from a careful examination of potential service opportunities is a digital library effort unlike anything that we have seen before. By mobilizing the creative talents and expertise of the leading institutions in this arena, the Aquifer collaboration stands to capitalize on collaboration in radically new ways. By organizing this effort under the umbrella of the Digital Library Federation, the project will be able to accomplish a long-sought dream, a distributed and open digital library composed of the best features that can be contributed by the leaders of the field.

Background

The Aquifer Services Working Group was chartered on February 28, 2005 during the third and final meeting of the Aquifer Prototype Group at Stanford University, California. The group's charter (after subsequent amendments) is comprised as follows from a charge and a set of immediate tasks:

WORKING GROUP CHARGE

The Services Working Group recommends service specification and level policies to the DLF Aquifer Implementation Group. The working group develops functional specifications for DLF Aquifer services, working closely with scholars. The Services Working Group is also responsible for designing processes to measure DLF Aquifer usability.

IMMEDIATE TASKS

- 1) Gather information about existing user studies and review.
- 2) Poll DLF Aquifer participant libraries and DLF for additional user studies information or findings from collaboration with scholars.
- 3) Develop a set of DLF Aquifer use cases for Phase I (assuming Dublin Core metadata).
- 4) Develop functional specifications for Phase I DLF Aquifer services.

SERVICES WORKING GROUP MEMBERS

- Martin Halbert , Working Group Chair (Director for Library Systems, Emory University)
- John Butler (Director, Digital Library Development Laboratory, University of Minnesota Libraries)
- Laine Farley (Director, Digital Library Services, California Digital Library)
- Michael Furlough (Digital Research & Instructional Services, University of Virginia Library)
- Beth Sandore (Associate University Librarian for Information Technology Planning and Policy, University of Illinois at Urbana Champaign)
- John Walsh (Associate Director for Projects and Services, Digital Library Program, Indiana University)

Analysis of Relevant DLF Activities

SURVEY OF DLF AQUIFER INSTITUTIONS

The Services Working Group has undertaken an initial survey of Aquifer institutions, to begin to assess what services are desired by end users and institutions and how institutions see Aquifer potentially meeting these needs. Surveying will continue through the summer to include all DLF institutions; at the time of this writing, Aquifer institutions are still in the process of scheduling interviews with the Working Group and responding to the survey. Initial responses are summarized below. As many results as are available will be incorporated into the final document prepared for the Aquifer meeting planned for June 28, 2005. A subsequent document will summarize all survey results.

Methodology

The group contacted those Aquifer institutions not already represented by members of the group. Contact persons at these institutions were given a survey and were asked to provide responses to the questions and then discuss the responses during a conference call with members of the Services Working Group. Notes taken during the conference call were sent to the interviewee for review, who made any necessary corrections. See Appendix B for an example of a completed survey.

Initial Assumptions

In creating this survey, we assumed some level of institutional awareness of digital-collection use, whether through mere observation of user behavior or through more systematic assessment of collection use and user needs.

In our discussions with contact persons, it became clear that we were also assuming that users could articulate what new services and functionalities would be useful to them in their work. It appears that users rarely analyze what new functionality they desire or need, if they indeed even reflect much about their specific service needs. Some researchers may have sophisticated searching needs which require them to be more aware of what features support or hinder their work; but most users have basic searching needs, which do not require them to be so attentive. Often users may conflate content with service, what some have referred to as “the Google effect.” In these situations, users’ narratives become particularly important for parsing what enhancements to services and/or collections they require for their work.

Initial Responses

Patron Use of Digital-Object Services & Collections

One institution discovered through user studies that not only do users tend to conflate service and content, but institutions would be well advised to consider how blurring these boundaries could support collection use.

Survey Response:

"There is reason to believe that drawing distinct boundaries between collections and services may not be appropriate. Based on our work to generate scenarios and use cases, it seems reasonable to submit that individuals are driven by use in addition to (instead of?) discovery. That is, it's more likely that someone will want to discover relevant content at the point of need while s/he works within a learning environment or a research collaboration space."

Additionally, users may be more inclined towards services that provide them with greater control over content: for instance, being able to define metadata for resources, using "wiki" technology. Services

Along more traditional lines, users may seek digital-object services that support research practices developed with print materials (i.e., searching by keyword, sorting results, etc.). But there is recognition that using digital-objects prompts uses and desires that traditional resources do not typically engender.

Interview note:

"Using digital resources is not like reading a book in the corner... when you're browsing the web, you find pieces you were not necessarily looking for, and you want to integrate those into various uses."

Library Assessment of Current Digital-Object Services

Institutions recognize resource limits (i.e., time, money, and staff) on their ability to develop digital-object services. User studies help to identify areas for improvement, but it is difficult to justify clearly desired services when such enhancements would serve only a small percentage of potential users. During one conference-call discussion, the observation was made that, while a service might benefit only a small group of researchers, those researchers might represent a sizable percentage of a digital collection's use.

Interview note:

"These people who have the specific requests are frequently the ones who come back a lot. ... Difficult to justify developing this type of tool for just a few people. Raises the question of whether we should be developing services and tools for these folks, since they would likely bring their students in to use the types of tools we developed for their work."

Aquifer is seen as having potential to pool resources for developing new services and to provide test beds for these services.

The following are services that surveyed institutions identified as desirable or needed for use with their own digital collections, and that they have not developed but would (given the time and resources):

- Registry of digital collection
- Cross-institutional searching of other digital collections
- Automated citation creation and plug-in with EndNote, Procite, or other bibliographic citation software
- Better integration of digital collection with OPAC
- Integration of multiple repositories and applications, using a common framework

USER STUDIES CONDUCTED BY DLF MEMBERS

Many DLF institutions have already undertaken studies of users and analysis of what opportunities and needs for new digital library services exist. We have sought out studies of this kind that seem particularly relevant to the Aquifer effort and have reviewed their findings. Studies we have so far noted and consulted include the following.

OAI Usability and K-12 Teachers

Experiences of Educators Using a Portal of Aggregated Metadata. Shreeves, Sarah, and Christina Kirkham. (2004, September 9). *Journal of Digital Information* 5(3), Article No. 290.

Abstract: The University of Illinois at Urbana-Champaign Open Archives Initiative Metadata Harvesting Project sought to test the viability of a search portal containing aggregated metadata for cultural heritage resources harvested using the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH). Metadata was collected from 39 providers, including museums, archives, libraries, historical societies, consortiums, and digital libraries. Some resources existed in digital formats, such as .JPG images. Other resources were analog objects and were represented digitally through the metadata. The paper documents a pilot user test with a small group of K-12 teachers-in-training. The users were asked to use the portal to locate primary source materials for use in the classroom. The results highlight the challenges posed by aggregations of heterogeneous metadata for both users and service providers. Areas for further investigation and approaches for more in-depth studies are suggested.

Semantic Clustering

MetaCombine Project Interim Report on Experiments at Emory in Combined Searching, Automated Organization of OAI and Web Resources, and New Forms of Scholarly Communication. Halbert, Martin, Aaron Krowne, and Katherine Skinner. (2004, September). MetaScholar Initiative Report, Emory University, Atlanta, Georgia.

Abstract: Conducted by the MetaScholar Initiative at Emory University, the MetaCombine project found that there are many effective techniques for semantically clustering (and thereby organizing harvested material for browsing) both metadata aggregated through the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) and web pages aggregated through web crawls. An important basic finding is that such semantic clustering techniques fall into two broad, complementary categories: 1) clustering uncataloged information to identify underlying patterns in order to create new classification schemes, and 2) classifying (or understood another way, assigning metadata to) uncataloged information once one or more classification schemes have been selected or

derived. The project has developed several effective systems for combined searching of harvested metadata and web pages. A basic finding related to the creation of such a combined function service is that it must include a technology called focused crawling, which provides the capability of selectively crawling only those websites relevant to the subject domain under consideration. Without focused crawling, the quality of web content aggregated for scholarly studies is greatly compromised. The project has also successfully produced a range of visualization tools and techniques that graphically display the results of semantic clustering systems. These tools are being vetted to focus groups of scholars. Early results have shown that scholars see value in the visualization techniques, and are also able to quickly suggest new scenarios for using the tools to conceptualize bodies of information.

Digital Resources in Undergraduate Education

Higher Education in the Digital Age Project. Center for Studies in Higher Education, University of California, Berkeley; Center for Information Technology Research in the Interest of Society (CITRIS), and the California Digital Library (CDL). (2003-2005). <http://digitalresourcestudy.berkeley.edu/>.

Abstract: The Center for Studies in Higher Education has received grants from the Andrew W. Mellon Foundation and the William and Flora Hewlett Foundation to investigate the use of digital resources in humanities and social science undergraduate education. Additional funding has been provided by the Center for Information Technology Research in the Interest of Society (CITRIS), the Hewlett-Packard Company, the California Digital Library (CDL), and the Vice Chancellor of Research, UC Berkeley. The completion date is July, 2005. This project will map the universe of digital resources available to undergraduate educators in the humanities and social sciences (H/SS), and examine how understanding use and users can benefit the integration of those resources into undergraduate teaching. The definition of digital resources is intentionally broad and includes rich media objects (e.g., maps, video, images, etc), as well as text. These digital resources may reside in or outside of digital libraries, and include those developed by individual scholars as well as other entities. Methods include discussion groups and surveys of faculty, graduate students, librarians, site owners and educational technology professionals. The project also employs online surveys and transaction log analysis on select local collections.

American West Project User Studies

Documenting the American West User Interviews – Final Report. California Digital Library (CDL). (2004, July 1). http://www.cdlib.org/inside/projects/amwest/americanwest_assessmentfindings.pdf.

Abstract: This report details the findings from a series of interviews with faculty, graduate students, librarians, and K-12 teachers, which attempted to assess what users needed and expected from online research tools and how they used these resources. Users in this study expressed a strong desire for primary source materials, interactive features in the digital collections (i.e., dynamic maps and narrative slideshows), multiple searching options (minimum of keyword, but also searching by format, date, location, and collection), citation management tools, publishing tools, and geographically based views of collections. Based on these findings, the project participants recommend the development of advanced searching and filtering tools, tools for personalizing collections, more features that promote interaction with the collection, and more contribution of materials by users and institutions.

2003-2004 Progress Report on “New Frontiers in the Digital Library: Social and Ecological Diversity of the American West.” California Digital Library (CDL). (2004). <http://www.cdlib.org/inside/projects/amwest/progress/>.

Abstract: Summary of mid-point findings of the California Digital Library’s project to assemble an American West virtual collection using resources from major research institutions. Highlights of key findings to date include: the value of curated, bounded, and annotated collections to users; identification of strengths and gaps in current collections, and strategies for filling these gaps; and development of functional requirements for tools in the areas of basic infrastructure development, collection building, curating, access, and customization. By projects’ end, an extensive virtual collection will be assembled, accompanied by tools specifically geared towards supporting digital library development, maintenance, and integration with other online learning environments. The complete report is located at http://www.cdlib.org/inside/projects/amwest/amwest_year1_report.pdf.

OTHER RELEVANT STUDIES AND RESOURCES

Web Design for Cultural Heritage Materials Presentation

Museums and the Web 2005 – Best of the Web Awards. Archives and Museum Informatics. <http://www.archimuse.com/mw2005/best/index.html>.

Abstract: The most recent winners in Museums and the Web’s annual award series, in which museum professionals recognize outstanding web designs for cultural heritage materials presentation. Award categories include [On-line Exhibition](#); [E-Services](#); [Innovative or Experimental Application](#); [Museum Professional’s Site](#); [Research Site](#); [Educational Use](#) and [Overall Museum Web Site](#), selected by the judges from all of the [sites nominated](#). These sites provide a variety of thoughtful examples for presenting and organizing information to users, including novel modes of browsing using place and time, and user driven views of collections.

Digital Resources in Humanities and Social Sciences

Scholarly Work in the Humanities and the Evolving Information Environment. Brockman, William S., Laura Neumann, Carole L. Palmer, and Tonyia J. Tidline. (2001, December). Council on Library and Information Resources. <http://www.clir.org/pubs/reports/pub104/contents.html>.

Abstract: The Scholarly Work in the Humanities Project began in 1999 to examine in detail how humanists work, how they are integrating technology into their work, and how future technologies might offer new opportunities in line with the goals of humanities research. The project was based on the premise that future development of research libraries should be informed by the actual practices and needs of working scholars and that it should take into account the value and impact of the technologies that they have adopted thus far. Decisions about how to build collections and services in research libraries should be more responsive to the disciplines that have historically depended on library and archival resources, and they should take into account the many types of resources and activities involved in the scholarly process. This report provides a foundation for developing user-based criteria for guiding digital library development by articulating what scholars do in the course of their research and how they depend on information to follow their paths of inquiry and write new texts. Through the analysis of scholars’ practices we can conceptualize the type of information environment that would

best support their activities and begin to clarify priorities for the development of rich information environments that are responsive to the context of scholarly work.

E-Resources for Research in the Humanities and Social Sciences; A British Academy Policy Review. The British Academy: The National Academy for the Humanities and Social Sciences. (April 2005). <http://www.britac.ac.uk/reports/eresources/>.

Abstract: In an effort to discover what impact e-resources and e-access are having on research in the humanities and social sciences (HSS), both in their own right and through their interaction with non-e resources, The British Academy consulted relevant national institutions and organizations and individual researchers in all areas of HSS to establish the present state of e-resources and their use, and to identify future needs. This review reports on what the organizations are doing to support resource provision, and on how researchers are actually working with resources, through an examination of the role of technological factors, of resource properties, access means, and holding requirements, and of publication and rights choices. The study finds that HSS researchers are proactive and enterprising in their use of e-resources, even where much of their work is with non-e resources; are conducting their research with as much e-sophistication as workers in science, technology and medicine; are developing HSS research in new ways; and are reaching across subject boundaries within and outside HSS. Much of this gain comes from using generic resources, notably email and Web engines. But much comes from HSS-specific resources, making it clear that e-resources are vital for HSS research health. Findings also indicate that the United Kingdom's (UK) resource provision is ad hoc and fragmented, structural options are not being fully analyzed, and important policy issues are not being vigorously addressed. More strategic, coordinated and well-targeted action is needed—grounded in researchers' actual, not deemed, requirements—so e-resources and access to them are designed, from the beginning, for researchers' use.

NSF SBE-CISE Workshop on Cyberinfrastructure and the Social Sciences. Berman, Francine, and Henry Brady. (2005). <http://www.sdsc.edu/sbe>.

Abstract: The National Science Foundation funded the SBE/CISE Workshop on "Cyberinfrastructure for the Social and Behavioral Sciences" in recognition of NSF's role in enabling, promoting, and supporting science and engineering research and education. The workshop was intended to help identify the SBE sciences' needs for infrastructure, their potential for helping CISE develop this infrastructure for engineering and all the sciences, and their capacity for assessing the societal impacts of Cyberinfrastructure. The report provides eight summary recommendations, including these first three: 1) Develop and deploy enabling data-oriented Cyberinfrastructure targeted to the social and behavioral sciences; 2) Develop and deploy targeted toolkits, virtual, and computational environments for facilitating social and behavioral science research; and 3) Instrument and design technologies to gather and provide key data for social scientists. Conversely, utilize human and computer interaction data to instrument and design Cyberinfrastructure technologies.

Aquifer Use Cases

Cockburn (1995) defines a use case as “a collection of possible sequences of interactions between the system under discussion and its external actors, related to a particular goal.” Beaudry (1996) points to the importance of use cases for clarifying what features of a system or service are required by its users. From conversations among group members and with other Aquifer institutions, we concluded that use cases offered the best approach to describing scholars’ experiences with digital collections and identifying functional requirements for creating services.

The goal-oriented use cases outlined below present the persons and conditions essential for achieving each stated goal, the factors which could interfere with successful achievement of goals, scenarios illustrating how the use case might appear in reality, and how the use case connects to Aquifer project phases. The use cases are divided into two groups, “End-User Use Cases” or “Institutional Use Cases,” according to who is most immediately involved in the action described.

These use cases should be considered as an initial set of targets for the Aquifer Implementation Group to consider; there are clearly many more potential services that can be envisioned. After reflecting on the survey responses and reports that we have so far considered, we feel that these are very strong candidates for Aquifer services, offering many benefits and flexible combinations.

END-USER USE CASES

Researchers include anyone who conducts research, and may include faculty, graduate students, independent learners, and undergraduates.

We devoted special conceptual focus to researchers that are members of Aquifer institutions. While Aquifer content may be discovered and accessed through a central interface, we assume that ultimately end-users will discover and use Aquifer content through the interfaces, systems and tools that their local research library provides in ways that seamlessly integrate this content with other content developed, managed or licensed by those institutions. This deep sharing will require the elaboration of a corresponding set of institutional services and use cases.

Use Case #1. Finding Research Materials

Goal: The most basic activity of researchers is seeking out materials that they can use for research activities. These materials may be items that they are already familiar with (“I remember I used that archival collection at our library last year”) or may be associated with topics they are interested in (“I wonder if there is anything out there on this subject”).

Primary Actors: Anyone who conducts research; may include faculty, graduate students, independent learners, and undergraduates.

Secondary Actors: None. While there are many other actors behind the scenes, none are required to provide assistance to the primary actors for them to accomplish their goal. In fact, not requiring mediation is a key requirement for this scenario and several of those listed below.

Aquifer Phase: This use case will be relevant to all three phases of the Aquifer project; however it will likely take different forms as the infrastructure of Aquifer progresses. In the first phase of Aquifer, researchers will be searching harvested metadata; in subsequent phases they will presumably be searching more than this.

Preconditions: There must be a searchable body of information assembled, metadata, full text, etc. For Libraries using Aquifer services, it may be important that materials available through Aquifer can be searched alongside their local and licensed content through their own interfaces, while for Aquifer contributors, appropriate attribution will be important.

Scenarios: Researchers come to a search interface after identifying an information need, search using either simple or advanced interfaces, and select the items they are most interested in. Searching is likely to be an iterative process in many cases, so users may be expected to loop through the procedure several times.

Success Conditions: The user comes away from the search process with one or more items that they consider valuable for research activities.

Failure Conditions: The user does not identify any useful items during the searching process. There are many potential ways this might happen. No items in the database may fit their criteria of interest. The interface might not work, or might be too confusing for them to effectively use. Results of searches might be accurate, but incomprehensible.

Comments: This is one of the most foundational activities the Aquifer will be called upon for, and is presumably a use case that we now understand well. It should therefore work well, and we should make sure that the central interface is appealing and effective. We also need to devote thought to what kinds of tools and facilities will lend support to this function in local Aquifer implementations. Such distributed search functions may take many forms potentially, ranging from search engines to web services. Facilities supporting this goal may in turn rely on infrastructure associated with inter-institutional use cases.

Use Case #2. Browsing Unfamiliar Research Archives

Goal: Another basic activity of researchers is the exploration of unfamiliar collections comprising a digital library. This is similar to the previous use case, but is different in that users may have a less specific target in mind and are seeking to evaluate what a collection contains.

Primary Actors: Anyone who conducts research; may include faculty, graduate students, independent learners, and undergraduates.

Secondary Actors: None.

Aquifer Phase: This use case will be relevant to all three phases of the Aquifer project; however it will likely take different forms as the infrastructure of Aquifer progresses. In the first phase of Aquifer, the browse capability may be relatively primitive, since researchers will be searching unqualified Dublin Core metadata, and may only be able to browse at the collection level. In subsequent phases they may be browsing information that has been classified by means of various controlled vocabulary structures.

Preconditions: There must be a body of information which can be browsed through some set of entry points. For Libraries using Aquifer services, it may be important that

materials available through Aquifer can be browsed along with their local and other licensed content available through their own interfaces, while for Aquifer contributors, appropriate attribution will be important.

Scenario: Researchers come to a browse interface after identifying some collection or collections they are interested in exploring, browse using various potential interfaces, and select the items or ranges of items they are most interested in.

Success Conditions: The user comes away from the browse process with a better understanding of the materials in the digital library, with an eye to evaluating the collections for subsequent research activities.

Failure Conditions: The user does not understand the structure of the digital library, and does not identify any useful items during the browsing process. There are many potential ways this might happen. The interface might not work, or might be too confusing for them to effectively navigate. The interface may work and be navigable, but they may not think the browsing structure is useful or coherent.

Comments: Effective subject browsing is currently something that is difficult to do with harvested metadata. It may be that this use case must wait for the second phase of Aquifer or later to be realized. Like other use cases, this one may take several forms (both centralized and distributed), and the distributed functions may rely on infrastructure associated with inter-institutional use cases.

Use Case #3. Discovering New Content through Alerting Services

Goal: Keeping abreast of newly published content and becoming aware of relevant content contained in newly-available services are universal needs among researchers across all disciplines. The goal is to provide timely, systematic, and low-effort mechanisms to deliver pertinent content to researchers based on personally-defined interest profiles.

Primary Actors: Anyone who conducts research; may include faculty, graduate students, independent learners, and undergraduates.

Secondary Actors: Few to none. While there are many other actors behind the scenes, none are *required* to provide assistance to the primary actors for them to accomplish their goal. Some researchers may opt for assistance in a higher level tuning of interest profiles by professional information services staff. However, this service is fundamentally unmediated.

Aquifer Phase: This use case is valid for any phase of Aquifer. However, this service will likely be considered as an added-value component that receives development attention when foundational services are deployed and stabilized.

Preconditions: There must be a searchable body of information assembled, metadata, full text, etc. There must also be capacity to create/edit/store/delete personal accounts. Ideally, for Libraries using Aquifer services, profiling/alert functions could target locally- and remotely-hosted content, including licensed resources and web search indexes. The profile-alert mechanism may live within or outside of Aquifer services. This architectural flexibility is important so as not to duplicate existing alert mechanisms that the user is already operating.

Scenarios: A researcher comes to rely on three to four growing databases providing a combination of text and image objects that hold potential as key primary sources to her research. Through experimentation with searching these databases, the researcher has found satisfaction with a query that finds the right balance of recall and precision. She records the query offline and commits to running it on a monthly basis to see if the query retrieves new results from the database updates. With the newly introduced profile-alert service, the researcher establishes a personal account, submits the query, the database targets to be hit by the query, and the frequency with which the profile should be run against the targets. The researcher also has the option of selecting between e-mail or RSS notification methods for alerting her to newly discovered records. The researcher may tune the profile and any of its parameters at any time.

Success Conditions: The researcher develops a trusted relationship with the alert service as one bridge of ongoing discovery to newly published content related to his or her research interests.

Failure Conditions: The user's profile is ineffective in its query construction or target selection, which may either result in generating "noise" or, possibly, no results, in which the user may ignore or forget about the service.

Comments: Building personalization options into services holds high-value potential for information discovery and research processes. While these services need to be simple and intuitive for users to manage, they are not necessarily simple services to construct on the backend. The information environment is, itself, a challenge of complexity. Providing account-level services in the context of that complexity adds another dimension to the challenge. Providing personalization options would likely require a significant up-front investment, but may deliver considerable benefit.

Use Case #4. Collecting Items in a Persistent Exportable Basket

Goal: Users often want to record, collect, and track groups of items that they find in a large digital library, in the same way that they collect bookmarks of web pages while navigating the Web. When they return to the digital library, they want the collection of records to persist so that they can refer to them and edit the list over time.

Primary Actors: Researchers who return to the Aquifer digital library repeatedly.

Secondary Actors: None.

Aquifer Phase: This use case is valid for any phase of Aquifer.

Preconditions: The items found in Aquifer must have some kind of persistent identifiers, even if these are only internal to Aquifer. A Library using Aquifer services to integrate available content into their own collections must provide such a basket capability to their users through their own discovery system.

Scenario: Users encounter items while using Aquifer interfaces and indicate that they wish to persistently track these items. When users return to Aquifer, they are able to easily access and continue to edit (through additions and subtractions) items in their persistent basket. Variations: Users are able to restrict or grant basket access to other users. Baskets may in turn become items that can be subject to all the other use cases (i.e. they may be searched, browsed, annotated, etc.). The basket may be able to export or save its contents in various formats, such as Endnote citations.

Success Conditions: Researchers routinely use the basket feature, and report that it enhances the value of Aquifer to them.

Failure Conditions: Researchers do not use the basket feature, whether because the interface is confusing, the feature is unreliable, or they simply don't understand the point of the feature.

Comments: Users benefit from baskets in many websites, especially online commerce sites. Like other use cases, this one may take several forms (both centralized and distributed), and the distributed functions may rely on infrastructure associated with inter-institutional use cases.

Use Case #5. Searching Full Text

Goal: Users often find it valuable to conduct full-text searches through the contents of digital collections, rather than just metadata associated with the item.

Primary Actors: Researchers.

Secondary Actors: None.

Aquifer Phase: This may or may not be possible during the first phase of Aquifer, depending on whether or not full text access is available.

Preconditions: The full text of a body of relevant items must be aggregated and indexed. This may require other use cases to be in place, such as the focused crawling use case.

Scenario: Users enter keywords in a full-text search interface and are able to effectively identify items of interest to them in the results.

Success Conditions: Researchers are able to use the interface effectively, understand what body of materials they are searching, and feel that the full-text search capability adds value to their use of the digital library.

Failure Conditions: Researchers cannot understand how to use the full-text search interface, or do not feel it adds value whether because of limited coverage or clumsy navigation or display of full-text information.

Comments: While this use case is optional in the first phase of Aquifer, it will be an essential aspect of the third project phase to enable deep sharing of information content. Like other use cases, this one may take several forms (both centralized and distributed). We will need to explore how this function can best be fostered in distributed, local applications.

Use Case #6. Annotate / Comment

Goal: A central activity of scholars is the commentary and annotation of research materials. The capability of a user community to jointly share notes on items of interest is an important mechanism for users to voice opinions and communicate with peers. This capability also adds value to the underlying records through additional context.

Primary Actors: Researchers who use the items found in the Aquifer DL.

Secondary Actors: None.

Aquifer Phase: This scenario could be implemented in any phase of the Aquifer project.

Preconditions: There must be persistent records in the database, and a mechanism for persistent annotation must be in place. If the annotations are to be used for metadata enhancement, this may require that other use cases be in place, such as metadata remediation or metadata enhancement.

Scenario: After accessing records in the database, users will enter annotations in the form of comments on the items and links to related information. Such annotations would allow for the development of new collection definitions/content clusters for browsing and could be used to enhance existing metadata. Variations: Users will be able to comment on comments, producing threaded discussions on particular collections or items.

Success Conditions: Researchers will routinely use the annotation system and report that it adds value to their experience of the Aquifer.

Failure Conditions: The annotation feature will be rarely used and no significant base of commentary will accumulate.

Comments: Annotation and ranking mechanisms are regularly identified as one of the most valuable features in Amazon and other portals. The hardest part of this function is jump starting it, getting a critical mass of users to use it enough that it develops some momentum. Like other use cases, this one may take several forms (both centralized and distributed), and the distributed functions may rely on infrastructure associated with inter-institutional use cases.

Use Case #7. Accessing Content Seamlessly through Course Management Systems

Goal: Much of the academic work students do online, including accessing primary materials such as a reading list and bibliography, is organized through their institution's course management system. Instructors should be able to gather, create and share digital content discovered via Aquifer services for direct access through the course management system, and students should be able to access the range of functionality appropriate for that content through that system. Such "packages" developed by the instructor should be able to be re-used from semester to semester.

Primary Actors: Students, instructors.

Secondary Actors: None.

Aquifer Phase: This scenario would be valid for the later phases of Aquifer implementation.

Preconditions: This scenario will likely require the development of at least Use Cases # 1, 2 and 3 above, as well as other institutional use case yet to be defined. It also assumes that a CMS is open and able to incorporate complex digital objects, including datastreams and their corresponding functionality. It is reasonable to assume that some preconditions for the success of this scenario are outside of the scope of DLF Aquifer.

Scenario: While developing her syllabus and course materials, an instructor identifies a set of digital resources available through Aquifer (a set of images and transcriptions of

manuscript items). She is able to “pull” the content and its corresponding functionality (such as multiple zoom views, annotation/commentary, slide shows, etc) into the area of her institution’s CMS reserved for her class, and there organize it in the way that best suits the needs of her class (slide shows, virtual exhibitions, etc). Student later access these materials to do their coursework without leaving the CMS. When the instructor teaches the course the following year, she can reuse or modify the collection of content.

Success Conditions: Instructor and student can use the CMS to access a richer set of digital content for instructional work.

Failure Conditions: The packaging ability will be limited. Instructors instead continue to develop these content packages themselves through PowerPoint, web pages, or other presentation tools, referring their students outside of the CMS to other tools or resources that manage the digital content needed for the class.

Comments: The Flecker & McClean study “Digital Library Content and Course Management Systems: Issues of Interoperation” provides many more specific scenarios for this type of activity.

INSTITUTIONAL USE CASES

There are many inter-institutional use cases that can be imagined. The following are meant to be some representative real-world examples that Aquifer institutions could take on, which would be immediately useful and which would enable experimentation with distributed DL functions for the Aquifer.

Use Case #8. Metadata Harvesting

Goal: Several DLF institutions have extensively explored and documented the inter-institutional activity of metadata harvesting. The basic aim of such efforts is to quickly assemble a union database of records contributed from many digital libraries. Such union databases have the obvious advantage of offering end-users access to content from many different institutional collections, usually through search services as described in Use Case #1 above. The creation of such union databases is enabled through inter-institutional collaboration using the OAI Protocol for Metadata Harvesting.

Primary Actors: DL staff who operate metadata harvesting systems in connection with the Aquifer project.

Secondary Actors: DL staff who have installed and maintain OAI data providers for content relevant to Aquifer efforts. These DL staff will frequently be at different institutions than the staff operating the harvesting systems.

Aquifer Phase: This use case is relevant to all subsequent phases of the Aquifer project.

Preconditions: OAI data providers for content relevant to Aquifer efforts.

Scenario: The Aquifer project will likely employ metadata harvesting systems in many related scenarios. The most straightforward scenario is the case of the central Aquifer union database, a metadata harvester and associated indexing system which is currently being built as part of IMLS-funded work. In this scenario, Aquifer institutions will maintain

OAI data providers and the central metadata harvester will aggregate records contributed by the providers. This will enable the foundational capability described in Use Case #1 above. Variations: There are number of additional scenarios in which metadata harvesting will likely be useful. Many localized interfaces to segments of the Aquifer content are anticipated in the course of the project, all of which will have to be populated through some means, with the OAI-PMH being the most likely mechanism. The various metadata modification use cases outlined below will require some means of metadata transmission, again with OAI-PMH being the most likely candidate.

Success Conditions: Metadata harvesting services must function reliably and be used by DLF institutions. Data providers must be created according to specifications laid out by the Metadata Working Group. It may also be that this and other services would be functions that Aquifer would like to offer more broadly to non-DLF members.

Failure Conditions: The services either do not work reliably or are too difficult for effective use by others.

Comments: Aquifer metadata harvesting functions will likely be a major area for immediate work in the project, especially since the DLF IMLS OAI project is already underway. There are many opportunities here for experimentation and innovation. Scenarios for re-harvesting of previously harvested records will be an area to explore, with potential benefits associated with iterative harvesting and processing functions as described below. Provenance of records will be an important issue to understand in these scenarios. Will we offer a central metadata provider repository in addition to a search service, ala the NSDL? There are many possibilities to examine.

Use Case #9. Metadata Remediation

Goal: One of the most frequently cited problems in metadata harvesting is the variation in metadata that is aggregated from different sources. Most operations that have undertaken metadata harvesting have engaged in various forms of metadata remediation or normalization to make heterogeneous metadata more consistent for searching and browsing functions. Aquifer institutions would like to have access to normalized metadata associated with digital content from other institutions. They would like to do this in support of many different local scenarios, for example, to create search/browse functions on locally selected subsets of such content, locally developed subject mappings, etc.

Primary Actors: DL staff who are working on particular local implementations of Aquifer services.

Secondary Actors: DL staff running Aquifer infrastructure services. These DL staff may or may not be at the same institution as the primary actors.

Aquifer Phase: This scenario is likely associated with the second and third Aquifer phases; however, this will require analysis during the first phase to understand what shape this service will take in subsequent phases.

Preconditions: A critical mass of Aquifer institutions must have set up OAI data providers.

Scenario: Some Aquifer institutions will operate services for metadata remediation, possibly via the W3C web services framework. Examples of such remediation services might include date and format normalization. Other institutions seeking to remediate

metadata would be able to transmit a stream of unqualified Dublin Core records to the web service and receive back a stream of records in which the target fields had been normalized. Variations: this could take the form of centralized or distributed services. The great opportunity represented by the Aquifer initiative is that distributed institutions will be able to assume responsibility for development and upkeep of such services, and that such distributed development will lead to healthy evolution of such services over time as different sites create new versions of such software.

Success Conditions: The metadata remediation services must function reliably and be used by DLF institutions. It may also be that this and other services would be functions that Aquifer would like to offer more broadly to non-DLF members.

Failure Conditions: The services either do not work reliably or are too difficult for effective use by others.

Comments: Similar to many Aquifer services, we will need to think through the responsibilities and commitments associated with such distributed scenarios. Much as the case with OAI metadata providers and harvesters, there will be the potential for problems in the interactions between institutions that operate such distributed infrastructure and those seeking to use it. But such distributed scenarios are archetypically Aquifer-like, in that they enable the DLF to act as a whole on a set of problems too large for any one institution.

Use Case #10. Metadata Enhancement

Goal: Another finding from the first generation of metadata harvesting services was that metadata that could be easily aggregated via the OAI-PMH was often inadequate in various regards, at least for specific portal applications. Libraries would like to have access to records with controlled vocabulary subject metadata, links to collection records.

Primary Actors: DL staff who are working on particular local implementations of Aquifer services.

Secondary Actors: DL staff running Aquifer infrastructure services. These DL staff may or may not be at the same institution as the primary actors.

Aquifer Phase: This scenario is likely associated with the second and third Aquifer phases; however, this will require analysis during the first phase to understand what shape this service will take in subsequent phases.

Preconditions: A critical mass of Aquifer institutions must have set up OAI data providers.

Scenario: Some Aquifer institutions will operate services for various kinds of focused metadata enhancement, possibly via the W3C web services framework. Examples of metadata enhancement services might include assignment of controlled vocabulary subjects and collection record links. Libraries seeking to enhance metadata would be able to transmit metadata to the service and receive back a set of enhanced records. Variations: this could take the form of centralized or distributed services. Again, the Aquifer initiative will be a valuable testbed for distributed DL services that can leverage expertise at individual institutions.

Success Conditions: Metadata enhancement services must function reliably and be used by DLF institutions. It may also be that this and other services would be functions that Aquifer would like to offer more broadly to non-DLF members.

Failure Conditions: The services either do not work reliably or are too difficult for effective use by others.

Comments: There are many types of metadata enhancement that might fruitfully be explored. It will be important to develop consistent mechanisms for such service interaction, we will be able to explore the options for such services in the Aquifer context.

Use Case #11. Taxonomy Assignment

Goal: After harvesting a large body of information on some targeted subject domain from multiple institutional sources, whether by means of the OAI-PMH or web crawling, libraries would like to be able to offer their users access to the material through a comprehensive, browsable taxonomy. Yet, content that can be aggregated through either the OAI-PMH or web crawling does not come with such a taxonomy of controlled headings. In fact, most efforts to date have only been able to create browse mechanisms using some variation on the source of the information, which is not what most DL users want. What libraries would like is either the capability to map a body of metadata or content to an existing organizational taxonomy (such as the Library of Congress Classification system), or to derive an ad hoc taxonomy for a body of material aggregated by other means. Once a taxonomy has been assigned to all of the constituent items aggregated, a browsing function can be established (as in Use Case #2 above).

Primary Actors: DL staff who are working on particular local implementations of Aquifer services.

Secondary Actors: DL staff running Aquifer infrastructure services. These DL staff may or may not be at the same institution as the primary actors.

Aquifer Phase: This scenario is likely associated with the second and third Aquifer phases; however, this will require analysis during the first phase to understand what shape this service will take in subsequent phases.

Preconditions: A critical mass of Aquifer institutions must have made their metadata or materials available for harvesting via OAI-PMH or web crawls.

Scenario: Some Aquifer institutions will operate services for taxonomy assignment, possibly via the W3C web services framework. Such services might work in a variety of ways, and might in turn make use of other services. Variations: this could take the form of centralized or distributed services. Again, the Aquifer initiative will be a valuable testbed for distributed DL services that can leverage expertise at individual institutions.

Success Conditions: Popular browsing services based on assigned taxonomies must be created in the Aquifer context.

Failure Conditions: Taxonomy assignment services lead nowhere and are not utilized.

Comments: This is a new area of DL technology and is still in its infancy. The Aquifer project has the potential to greatly benefit from this type of service, as well as leading the research and practice of such functions.

Use Case #12. Focused Crawling

Goal: Aquifer institutions may wish to perform focused web crawling of web-published materials to round out a collection for access or to ensure persistence of those materials. Such materials may not be available in any other way for integration with a library's collections (e.g., via OAI harvesting, via query options for metasearching such as Z39.50, XML gateways, SRU/SRW, etc.) Aquifer institutions will need an infrastructure to provide tools to those who initiate the crawls as well as to manage the results of crawls. Mechanisms for sharing the results of crawls with other institutions can reduce the need for redundant collecting.

One of the challenges in web crawling is how to cope with the resource-intensive nature of the activity that can consume time, storage, remote servers, and network bandwidth. The ability to control the parameters for crawling selected web sites is a key component of this service.

Primary Actors: DL staff or librarians serving as curators who are working on particular local implementations of Aquifer services. Curators make decisions about which sites to crawl and how to configure the crawl.

Secondary Actors: DL staff running Aquifer infrastructure services. These DL staff may or may not be at the same institution as the primary actors.

Aquifer Phase: This scenario is likely associated with the second and third Aquifer phases; however, this will require analysis during the first phase to understand what shape this service will take in subsequent phases.

Preconditions: Aquifer institutions must have an identified need to include content that is only available via web crawling.

Scenario: Aquifer institutions might provide tools for curators to crawl web sites including the following: capability to register candidate sites and initiate web crawls using seed list of URLs; specify crawl parameters (e.g., frequency, duration, maximum file size, number of subdomains to crawl, robot exclusions, etc.); monitor crawls in order to estimate completion time, change parameters; capture metadata during the crawl; assess crawl quality (e.g., logs of files/bytes, comparison to previous logs, random sampling, analysis of anomalies, validation of crawl parameters); manage and describe collections, crawls, and harvested nodes; and search and browse crawled content. The specific services or tools in each of these areas could be determined based on the relative importance of this service for Aquifer participants.

Success Conditions: The focused crawling services must function reliably and be used by DLF institutions. It may also be that this and other services would be functions that Aquifer would like to offer more broadly to non-DLF members.

Failure Conditions: The services either do not work reliably or are too difficult for effective use by others.

Comments: Unlike other Aquifer services, focused web crawling does not depend upon or benefit from explicit or implicit relationships with data providers.

Recommendations and Next Steps

While this report includes a substantive set of initial findings and use cases, it must still be understood as a first effort to analyze an immense topic: the next generation of digital library services in the DLF-Aquifer context. There are many more activities that obviously must be undertaken; we offer the following recommendations and next steps:

1. **Environmental Scan:** A more comprehensive environmental scan should be undertaken to inform the work of the Services Working Group. The survey of DLF institutions must be completed and analyzed to discern findings. We anticipate completion of the survey by August 31, 2005. A more complete canvas of other relevant reports on service needs and opportunities is also needed. We will continue to examine other reports that we identify, and which are drawn to our attention as relevant. The survey findings and canvas of reports will be combined to serve as an environmental scan. This environmental scan will be prepared and submitted by the Services Working Group by the time of the Fall 2005 DLF Forum, if not before.
2. **Coordination with other Working Groups:** The Services Working Group will take steps to coordinate its efforts with the other Aquifer groups. Concrete steps will include having the chairs of the other groups participate in several of the upcoming conference calls of the Services Working Group, as well as proposing that the Services chair participate in some conference calls of the other groups. Members of the Services Working Group will actively follow the contributions of the other groups on the Aquifer online collaboration site (Basecamp). Group members will actively seek to propose ad hoc discussions with other groups as needed. Reporting process and communication activities with the Implementation Group will be worked out. Finally, we will work to identify logical planning steps for coordination with the other groups.
3. **Functional Requirements:** The Services Working Group will develop a set of functional requirements for Aquifer Phase 1. The form of these requirements and the deadline for their completion will be identified in consultation with the other Aquifer groups. A target has been put forward of having working Aquifer prototypes ready by the time of the Spring 2006 Forum; we will have to act promptly to meet this timetable.
4. **Develop a Planning Schedule for Subsequent Phases:** In order to prepare for subsequent project phases, we will need to develop a common planning schedule for development of functional specifications for Phases 2 and 3. A planning schedule will be worked out in consultation with the other Aquifer groups.

Appendix A: Additional Sources Consulted

Cockburn, Alistair. "Structuring Use Cases with Goals." *Humans and Technology Technical Report 1995.01*. This report was later published in the "Journal of Object-Oriented Programming," in two parts, the Sep-Oct 1997 issue and the Nov-Dec 1997 issue. URL: <http://alistair.cockburn.us/crystal/articles/sucwg/structuringucswithgoals.htm>

Beaudry, Paul. "Using Use Cases." 1996. Online Computer Science Tutorial. URL: <http://www.csc.calpoly.edu/~dbutler/tutorials/winter96/use/>

Appendix B: Sample Response to Aquifer Institutional Survey

In an effort to determine what digital-object services are most desired and needed by Aquifer member institutions, the Aquifer Services Working Group is conducting this survey of library services. The following questions are intended as prompts for thinking about how users currently use digital collections, how that use might be improved, and what is needed in order to improve services. Questions in Part I focus on actual use of digital-object collections and services, particularly evidence that illustrates how different library patrons (faculty, students, and researchers) utilize digital-object collections and services. Questions in Part II focus on the library's own assessment of current digital-object services, particularly gaps in service and what is needed to bring about improvements.

Please look over these questions and provide answers to as many as possible that are relevant to your institution's situation.

Part I. PATRON USE OF DIGITAL-OBJECT SERVICES & COLLECTIONS

In this section, we are looking for evidence of patrons' use of digital-object services and collections. Preferably, this information should come from formally conducted user studies of your library populations (faculty, students, and researchers) and user comments on library services; anecdotal or observational evidence is also useful.

1. **How are digital object collections used?**

To the extent possible, please provide evidence for this use (i.e., statistics) and scenarios of typical use.

Our ideas on why and how people use the collections are not generally backed up with good evidence. We think that people use them for the same reasons they use print collections: seeking known items, reading longer passages or even entire books, and searching for keywords or names. Our systems are developed to support these activities.

2. **What digital-object services and functionalities do patrons think are necessary to support their work?**

Consider both the services identified as necessary and specific functionalities of those services which make them necessary.

As suggested above:

- seek known items by searching bibliographic data
- perform keyword or phrase searches
- limit searches by title, author, date
- sort results in various ways
- have the ability to page through works
- speed, speed, speed—nothing is more of a research buzzkill than waiting for the system/network/internet

3. **What digital-object services and/or functionalities are desired by these populations?**

More specifically, what digital-object services would patrons like to see added? And what functionalities would they like to see added to existing digital-object services? To the extent possible, please provide ideal-use scenarios that illustrate these desired enhancements.

I think people would prefer more content over added functionality, but that's just a hunch.

4. **If available, please provide a report of your library's most recently conducted assessment of services.**

Part II. LIBRARY ASSESSMENT OF CURRENT DIGITAL-OBJECT SERVICES

In this section, we are interested in the library's assessment of services: what services are needed or desirable, and what is necessary to develop those services? These could be *end-user services* or *institutional services* that support patrons' use of digital objects.

1. **What unmet service needs have you identified?**

These may be services, tools, or functionalities that could be added to existing services or tools. Also, consider whether these are gaps in *end-user services* or *institutional services*.

2. **What tools, services, or functionality have you *not* developed but *would* develop, given the time and resources?**

Consider whether these developments would be utilized directly by the end-user or by institutions in the service of the end-user.

- Registry
- cross-institutional searching
- automated citation creation and plug-in with EndNote, Procite or other bibliographic citation software
- better integration with OPAC

3. **What prevents you from accomplishing your service goals?**

Consider what factors constrain development of digital-object services (i.e., insufficiently trained personnel, outdated technology, budget allocations, etc.).

The usual: time and money. In addition, the lack of information and understanding about what users really want.

4. **What functional specifications for tools or services has your organization developed that you are willing to share with the Aquifer team, e.g., browse functionality, basket functionality, etc.?**

We have a mature digital library system with a great deal of functionality. Happy to talk about sharing any/all of it.