

The DLF Aquifer Initiative

Katherine Kott, DLF Aquifer Director

Association of Jewish Libraries

June 20, 2005

Oakland, CA

Description: DLF Aquifer emerged as the re-awakened strategic direction of the Distributed Open Digital Library initiative of the Digital Library Federation in May 2003. According to the original 1995 Digital Library Federation mission statement, the DLF was established to “bring together -- from across the nation and beyond -- digitized materials that will be made accessible to students, scholars, and citizens everywhere, and that document the building and dynamics of America's heritage and cultures.” DLF has progressed towards this strategic goal since its inception through support, coordination and participation in the development of prototypes, proofs of concept and test-beds that will form the foundation of DLF Aquifer. This presentation will focus on organizing for collaboration, leveraging existing collections and technical developments and defining the DLF Aquifer problem space.

Katherine Kott is the director of the DLF Aquifer Digital Library initiative. Her professional career has included experience in academic library systems, and in technical and public services. Most recently, Kott was the head of cataloging and metadata services at Stanford University, where she is based. Before coming to Stanford, she led the implementation services department at a major ILS vendor, coordinating the installation of systems at a wide range of libraries, including consortia.

I am pleased to have the opportunity to talk with you today about the Digital Library Federation and the DLF Aquifer initiative. To begin, I will provide some background information about the Digital Library Federation and how the DLF Aquifer initiative fits with other DLF activities. There are several collections of potential interest to this audience that are likely to be incorporated

into DLF Aquifer and I will highlight these. In addition, I will outline the “problem space” that DLF Aquifer is meant to address and discuss some tools and services the initiative is likely to incorporate. I hope we will have time for questions and comments. I would be especially interested to hear your perspectives on the ways in which a set of tools and services like DLF Aquifer would be useful in your library or ways that we might think differently about what we are doing.

The Digital Library Federation is a membership organization of research libraries dedicated to creating and supporting standards and best practices for the development and maintenance of digital libraries. Founded in 1995, the Digital Library Federation recently became a separate not for profit organization. Previously it operated under the auspices of the Council on Library and Information Resources. Although the DLF is now a separate organization, it is still viewed as the digital library program of the Council of Library and Information Resources and the DLF plans to remain in an office suite shared with CLIR in Washington, D.C. The Digital Library Federation is a lean organization. Besides me, there are three other DLF employees, the Executive Director, David Seaman, a program officer and an administrative assistant. Unlike RLG or OCLC or even the California Digital Library, DLF is not a service provider but relies on member libraries to host the functions it supports.

When the DLF was founded, the original charter included the vision of a distributed open digital library. Through a strategic planning process undertaken in 2003, DLF Aquifer emerged as this distributed open digital library of the Digital Library Federation. From 2003 until 2005, DLF Aquifer participant libraries made best efforts to define and undertake the initiative. This proved difficult as a fully distributed effort amongst people with other primary responsibilities. In fall 2004, the DLF advertised for an Aquifer director and in January 2005, I came on board in that capacity. I remain based at one of the participant libraries, Stanford University, where I was the head of cataloging and metadata services before joining the DLF.

My first task was to interview the people who had been involved in the effort to better understand goals and expectation for the project. Eleven DLF members had been actively engaged in creating a vision for the initiative and moving the project forward. Through the interview process, I determined that what most participants hoped to create was a test-bed of services geared towards meeting the needs of scholars but to be implemented within library environments rather than delivered as a product directly to the end user. The group had already settled on building the tools and services to operate with a carefully selected collection of digital material focused on American culture and life. With this information in hand, I drafted a business plan, which was approved by the participant libraries in February and by the DLF Board in May. For those of you who are interested in the details of the initiative, you can find the business plan on the DLF web site at <http://www.diglib.org/aquifer/AquiferBusinessPlan.pdf>. By the spring DLF Forum meeting in San Diego in April 2005, a twelfth library had joined the effort. Currently, the participants include the California Digital Library, Emory University, Indiana University, Johns Hopkins University, the Library of Congress, New York University, Stanford University, the University of Illinois, Urbana-Champaign, the University of Michigan, the University of Minnesota, the University of Tennessee, and the University of Virginia.

One of the trickiest elements of managing a project like DLF Aquifer is to create a scope that offers opportunities and challenges while defining tasks that can realistically be accomplished within the context of the resources the DLF member libraries are able to contribute or that we anticipate being able to support through outside funding. In creating the business plan, we defined the initiative with a focus on tools and services that will provide access to aggregated collections. DLF Aquifer will be a middleware layer, if you will. It will interoperate with repository solutions such as DSpace and Fedora and with content management systems that provide structured access to digital materials. DLF Aquifer will also interoperate with e-learning systems such as Sakai. As personal content management systems aimed at the scholar come along, we anticipate enabling DLF Aquifer interoperability with these as well. Later, I will show some examples of these kinds of systems in development. While the digital content that DLF Aquifer will draw from initially will be DLF member library collections, we anticipate that in future phases we will expand beyond the DLF to include other digital collections that can be made available persistently.

The DLF Aquifer project plan is divided into three development phases. DLF Aquifer working groups have created project proposals that will leverage existing collections and adaptively re-use, develop and test tools and services that have already been created by DLF Aquifer participants or other libraries with which the DLF has close ties.

In phase two, the existing tools will be further enhanced, and new tools and services will be developed. Collections that round out the DLF Aquifer scope will be sought and added. The third phase will implement “deep sharing” which will enable the delivery of the digital object from its home repository to the user’s desktop or local environment. In this phase, we also anticipate offering annotation tools and the capability to re-deposit modified objects into the repository of origin or perhaps an altogether different one.

Phase I of the project is underway, with a prototype demonstration target date of DLF Spring Forum in April 2006. Four DLF Aquifer working groups in the areas of collections, services, technology and metadata have just submitted sub-project proposals in their areas. Next week, all members of the DLF Aquifer working groups will meet face to face in Chicago to put the sub-projects into an overall project plan. The basic activities over the next several months will focus on identifying existing collections, harvesting metadata, and implementing existing tools and services for access and retrieval.

Several Digital Library Federation member library collections that will be incorporated into DLF Aquifer may be of interest to this group. The American Memory project at the Library of Congress has a number of rich collections focused on American culture. This American Variety Stage collection includes a sub-collection of Yiddish playscripts. American Memory also includes digitized historic motion pictures in a variety of formats. It is possible to download this early Edison film of a New York fish market to the desktop.

The California Digital Library hosts archival collections through the Online Archive of California. The structured metadata for these collections enables excellent finding capabilities. For example, it is easy to gather historic photographs of synagogues in San Francisco, even though these archival collections do not focus particularly on Jewish culture and life. Similarly, the Wright fiction collection at Indiana University is a broad collection of American literature, but the metadata applied to this collection makes it easy to identify material of potential interest in Jewish studies. Johns Hopkins University has digitized and kept intact a sheet music collection that once belonged to Lester S. Levy, a Baltimore straw-hat manufacturer, talented amateur musicologist and sheet music collector. While Levy’s collection encompasses all facets of American life, the indexing system he developed for his collection allows material pertinent to Jewish studies to be retrieved easily from the full collection.

One of the greatest challenges for the DLF Aquifer project will be to create a context for the federated collections that will meet users’ needs. Digital materials have typically been organized according to some kind of collection metaphor, either by subject or material type that makes sense within an institutional context or by, as in the Levy sheet music example, keeping material that was collected and organized by a donor intact. As these

varied collections are aggregated, it will be important to think about what metaphor for organizing makes the most sense. What are the research questions that DLF Aquifer will answer? As the collections are aggregated, what tools and services can be brought to bear that will improve access? How can the existing metadata be enhanced? Will it be possible to apply a DLF Aquifer taxonomy to the aggregated collections? What new developments might be appropriate to test? For example, how might semantic analysis apply to the DLF Aquifer resource pool? What if the expert who is using the material could add descriptive metadata as he or she does works with the digital objects?

An example of work that is in progress to create context for aggregated collections is the implementation of the Grokker software by Groxis in the Stanford environment. In the Stanford implementation, the Grokker software allows multiple targets to be searched simultaneously. Although in the current implementation, the result sets are not combined, seeing the results from more than one source represented graphically rather than in text list format is more intuitive for most.

Experience attempting to apply traditional cataloging processes to digital collections has demonstrated that these traditional processes do not scale to large bodies of digital material. Open source metadata creation tools that create structured subject access to massive quantities of content are in development at Emory University. In future phases of DLF Aquifer, such tools may allow discrete collections coming into the DLF Aquifer stream to have consistent metadata automatically generated for the full body of material. These mathematically based pre-processing tools use techniques called latent semantic analysis or latent semantic indexing.

Studies such as the scholar's panel sponsored by the Digital Library Federation point up the request by scholars for tools enabling digital objects to be annotated. Such annotations could include metadata enhancement like the addition of subject headings or names for newly identified subjects in photographs or to provide context for a specific purpose, such as inclusion of the material in a learning object for use in a course management system. As these capabilities are developed, DLF Aquifer will facilitate deep sharing and interoperability. Decisions about when to incorporate functionality within DLF Aquifer and when to use tools that have been developed in other contexts is still to be determined. For example, it may make sense to make adaptive re-use of the annotation capabilities being developed at UC Berkeley as part of the Scholar's Box initiative or to use the tools developed at Rice University for Connexions, described as a "content commons" for educational material, where learning objects can be placed in context.

The opportunities and challenges before us are great. One of the most interesting aspects of the project is to test capacity to support a truly networked initiative. There is no lead library for this project. It is a true collaboration. While collaboration enables results that are greater than the sum of the parts, it also requires altruism—decreasing one's own fitness to increase the fitness of another. For example, the participant libraries are devoting staff resources to this initiative when they could be applying their time to efforts that carry a high priority locally. DLF Aquifer, along with other Digital Library

Federation efforts also offers the promise of a reinvigorated role for research libraries in an “Amazoog¹” world.

For more information or for copies of the slides accompanying this presentation, contact Katherine Kott, DLF Aquifer Director by e-mail at: kkott@clir.org

¹ Term coined by Lorcan Dempsey Vice President of Research and OCLC Chief Strategist