The Berkeley Finding Aids Project:

Providing Access to Images Through SGML Encoded Text

Bernie Hurley, Director of Library Systems, UC Berkeley Library

The Berkeley Finding Aid Project is a collaborative endeavor to test the feasibility and desirability of developing an encoding standard for archive, museum, and library finding aids. Finding aids are documents used to describe, control, and provide access to collections of related materials. In the hierarchical structure of collection-level information access and navigation, finding aids reside between bibliographic records and the primary source materials. Bibliographic records lead to finding aids, and finding aids lead to primary source materials. The Project will involve two interrelated activities. The first task will be to create a prototype encoding standard for finding aids. This prototype standard will be in the form of a Standard Generalized Markup Language (ISO 8879) Document Type Definition (SGML DTD). Researchers at the University of California, Berkeley will develop the encoding standard in collaboration with leading experts in collection cataloging and processing, text encoding, system design, network communication, authority control, and text retrieval and navigation. Project participants will analyze the structure and function of representative finding aids. The basic elements occurring in finding aids will be isolated and their logical interrelationships defined. The DTD will then be developed based on the results of this analysis. Building a prototype database of finding aids is the second objective of the Project. Available hardware and software will be evaluated. Hardware and software will be selected to support the following basic tasks and functions:

1. create finding aids marked according to Project defined DTD;
2. create a database of finding aids on a SGML aware server;
3. link the Finding Aids to the related images;
4. provide SGML aware software for searching and browsing the finding aids and viewing the images over the Internet.

Client software must support display of a variety of graphic formats (TIFF, etc.). The client/server software should support a variety of search types: boolean keyword, word adjacency and proximity, and relevance ranking and feedback. The text viewing and navigation component of the client software should allow dynamic generation of an expandable table of contents adjacent to document text to supply context clues for reading comprehension and random, informed access to the text. Software should support hypermedia links between text and text and text and graphics. The finding aid database will serve two primary purposes. First, it will provide the encoding standard developers with computer application experience with which to refine and inform the development process.

Second, it will provide a means for end users to evaluate the utility and desirability of encoded finding aids that will enable them to provide new ideas and suggestions to the encoding standard developers. End users will include not only public users, but staff users as well. Optimally, while the test database server will reside in Berkeley, clients will be available at collaborating institutions. The Berkeley Finding Aid Project envisions an information future in which serious scholars and the casually curious alike can easily isolate the cultural treasures they seek. In this information future, information seekers follow clearly marked paths through library catalogs to finding aids and from finding aids to treasures in a multitude of computer and traditional formats and back.