Project Plan for Cross-Database Search Tool
January 2000 Release of the CDL Web Site

D R A F T

5/10/99; rev 6/8/99 lxf

After assessing discussions and themes at the recent Digital Library Forums, reviewing "Key Challenges" identified by the Tools and Services Working Group, and evaluating feedback from other sources, the CDL has decided to focus on a tool to facilitate simultaneous searching across multiple resources for the January 2000 release of the CDL web site.

Rationale:

A theme in a number of the discussions at the Digital Library Forums was the need for a simple search option to provide "one-stop shopping" as a solution to user confusion about where to begin searching, given the multitude of databases, catalogs, search engines, directories, web pages and other discovery tools. Faculty comments from the Editors Forums and other conversations with individual faculty members and library staff have also affirmed the desire for this type of tool. The interest at various UC libraries in Database Advisor, developed by UCSD, is further evidence that such a tool is needed.

This type of tool addresses one of the CDL's key functions which is to integrate resources. This path toward integration began with the Melvyl Union Catalog and Periodicals databases, and continued by linking the abstracting and indexing databases to serials holdings and then to full text of articles. The most recent examples include the integration of databases, electronic journals and OAC finding aids in the CDL Directory, and linking Catalog records to OAC finding aids.

The July 1999 release of the CDL web site will take another step toward integration by offering the following options:

- A single, simultaneous search across backfiles of CDL-hosted abstracting and indexing databases (such as MEDLINE/HealthStar, BIOSIS, and INSPEC).
- Links from external databases (such as Compendex, and SocAbstracts) to holdings in the California Periodicals database.
- Links from external abstracting and indexing databases to full text from licensed journal publishers (for example, MLA Bibliography linked to JSTOR titles).

As more disparate collections of resources are created that offer specialized search functions and other tools, it is increasingly important to provide an initial method of discovering relevant material regardless of format or location.

In addition, user expectations are being shaped by the web environment which offers a simple, apparently inclusive approach to discovering resources of all types. Although many faculty and some students recognize the limitations of web search engines, they also want similar features for searching high quality research materials.

Parallel developments:

While a tool to search across different resources is being developed and extended, other efforts will focus on better ways to integrate resources in the various UC databases. For example:

- *Melvyl Union Catalog*: The Union Catalog Technology Project (http://www.cdlib.org/libstaff/technology/projects/unioncat/) is investigating and testing various options for the future architecture of the union catalog. These studies are looking for more flexibility in configuring the Catalog to identify different formats, and possibly to extending the Catalog to include digital objects in the future.

- *Database of licensed digital resources:* The CDL Directory was created in part to provide ready access to licensed digital resources and finding aids which are currently not easily isolated in the union Catalog. The CDL is exploring what capabilities the Catalog needs in order to obviate the need for the Directory in the future. The Directory is serving as a testbed for features that might be desirable in the future union Catalog.
- *Database of open access resources*: The debate over how to provide access and control of these resources continues with experimentation on a number of fronts. It appears that some categories of these resources may be represented in the Catalog, while tools such as Infomine may continue to be used for other categories. Data migration between Infomine and the MARC format is being developed.

Because these other projects are in development and it is not yet clear what the "best" primary access tool will be or even whether there should be a primary tool, the cross-resource tool can provide convenience and coherence, at least at the first level of searching. This tool can provide integration even if repositories of resources change in character, mode of access, or become subsumed by other repositories. For example, the Directory may cease to exist if the future union catalog can provide the same functions. It is likely that specialized search tools will always be needed for more in-depth searching of various resources.

Proposal for the Search Tool:

The long term goal is to have this search tool work for all types of resources considered part of the digital library, but each resource must be analyzed for search options and protocols, filtering possibilities, subject groupings and other issues. Some resource categories are still in experimental or developmental stages. Therefore, the recommended priority for types of resources to be included is as follows:

Phase 1:

These resources have been identified for Phase I because they are stable and in production, they represent the core electronic resources that are part of the California Digital Library, and for the most part, they are known to support the technical underpinnings to enable the search tool. There may be some exceptions to the latter criterion for some of the external abstracting and indexing databases and reference tools, which will be evaluated on a case-by-case basis.

- Melvyl Catalog
- California Periodicals Database
- CDL-hosted abstracting and indexing databases
- External abstracting and indexing databases (including those licensed by individual campuses)
- CDL Directory
- OAC
- Infomine
- Reference tools (Encyclopedia Britannica, Gale Associations Unlimited, CIS, LION, etc.)

It may be desirable to include UC library catalogs, possibly as an option to retrieve on order and reserve materials.

Phase 2:

Further discussion will be needed to identify candidates for Phase 2, depending on the progress and experience gained in Phase 1. Some likely candidates include:

- Alexandria
- AMICO
- Government information (if there are sources other than the Catalog and Infomine)

The tool will build on the experience of Database Advisor, especially in the areas of search protocols, maintenance issues, and problems with specific resources. The DBA developers have agreed to share their

code and expertise, and will likely serve on the project team. However, the goal is to extend the tool beyond Database Advisor's capabilities as much as possible. Additional functions and phases will be specified by the project team.

The project team (assisted by the Tools and Services Working Group) will also evaluate possible commercial solutions, including Northern Light, and IAC/Gale's new TotalAccess product. Other issues include:

- Whether the tool should be hosted centrally or distributed?
- How it should be maintained as database producers constantly change their services?
- To what extent can campuses customize for their environment?
- What is the optimal level of granularity for searching across resources (i.e., what is the definition of a "resource")?

Staffing

The proposed *Project Team* could consist of the following:

Laine Farley, CDL (overall project management)
Sherell Holcomb, CDL (project management assistance)
Christy Hightower, UCSD (Database Advisor developer, database analysis)
Sherry Willhite, CDL (database analysis, functional specifications)
Brian Warling, CDL (web site design)
CDL/T programming staff (to be determined; probably Ken Weiss, Lynne Grigsby-Standfill, contractor or new P/A III)

The Project Team will coordinate and write functional specifications, coordinate review and campus input, do the programming, and ensure that the tool is integrated with the CDL web site and other CDL tools and services.

Project Consultants

UCB representative since UCB is using DBA now (e.g., Camille Wanat) (functional specifications)
UCSC representative since they are using DBA now (e.g., Ann Hubble) (functional specifications)
Margaret Mooney, UCR (Infomine coordinator)
Marsha Fanshier, UCSD (Database Advisor programmer)

The Project Consultants will participate as needed in developing functional specifications, advise on programming issues, provide information about specific resources to be searched, and provide expertise on implementation details.

Other Groups

Other advisory and special interest groups will be consulted as the project develops. These may include the Users Council, the Tools and Services Working Group, the Education Working Group, the Evaluation Liaisons, SOPAG, the UC consortium and bibliographer group chairs, the Technical Architecture and Standards Working Group, etc.

Ongoing staff resources:
1 FTE programmer
.25 FTE analyst

Process and Timeline:

Late June-early July:

Hold brainstorming session with Project Team, Project Consultants and interested members of the Tools and Services Working Group to develop initial list of functional requirements.

July:

Tools and Services Working Group members will visit campuses to discuss and refine functional requirements, build on other ideas, begin to prioritize functions. The group will also try to meet with the faculty members of the Universitywide Committee on Research Policy, a group that has expressed interest in the development of CDL tools and services.

July30:

Draft functional specifications, including phases and priorities; campus review via Users Council

September 3:

Final specifications for phase 1; programming begins

November 3:

Test system available to campus reviewers

November 19:

Feedback evaluated and changes defined; instructional activities initiated (to be determined by Education Working Group)

January 5:

Public release

Spring term:

Evaluation activities (to be determined; could include focus groups, online surveys, statistics collection, etc.)

Late spring:

Assessment of enhancements, later phases, if needed