The UNIVERSITY *of* CALIFORNIA LIBRARIES

FINAL REPORT: DECEMBER 2005

Bibliographic Services Task Force

Rethinking How We Provide Bibliographic Services for the University of California

Table of Contents

Executive Summary

Society is in the midst of learning how to "be" in the information age. The advent of computers and the inclusion of the Web in our work and private lives have pushed innovations and embraced information and access in ways we can hardly imagine. We are living in a complex and challenging digital landscape that changes constantly.

On the Library front, our bibliographic systems have not kept pace with this changing environment. The continuing proliferation of formats, tools, services, and technologies has upended how we arrange, retrieve, and present our holdings. Our users expect simplicity and immediate reward and [Amazon](), [Google](), and [iTunes]() are the standards against which we are judged. Our current systems pale beside them.

The current Library catalog is poorly designed for the tasks of finding, discovering, and selecting the growing set of resources available in our libraries. It is best at locating and obtaining a known item. For librarians and for our users, the catalog is only one option for accessing our collections. We offer a fragmented set of systems to search for published information (catalogs, A&I databases, full text journal sites, institutional repositories, etc) each with very different tools for identifying and obtaining materials. For the user, these distinctions are arbitrary.

Within Library workflows and systems too much effort is going into maintaining and integrating a fragmented infrastructure. We need to look seriously at opportunities to centralize and/or better coordinate services and data, while maintaining appropriate local control, as a way of reducing effort and complexity and of redirecting resources to focus on improving the user experience.

Books are not going away. Traditional information formats are, however, being used in combination with a multitude of new and evolving formats. It is our responsibility to assist our users in finding what they need without demanding that they acquire specialized knowledge or select among an array of "silo" systems whose distinctions seem arbitrary.

The famous sage Howard Cosell once said, "What's popular isn't always right. What's right isn't always popular." We suspect when it comes to the Internet and how it has simplified searching, what is popular is also right.

Below are listed the Bibliographic Services Task Force's core recommendations for actions we must undertake if we are to remain viable in the information marketplace.

I.   Enhancing Search and Retrieval

I.1   Provide users with direct access to item

I.1a: Have UC eLinks take you to a logical, default choice, with option to go back to the menu if you want a different option. (If there is a reliable full-text link that would be first choice. This assumes that in the majority of times, we could correctly anticipate what service the user would want.)

I.1b: Provide an "I-want-this" button that is present when the context warrants, with the goal of always offering a fulfillment option. No dead ends. Give the user an option to specify turnaround time; work behind the scenes to fulfill as well as we can.

I.2   Provide recommender features

I.2a: Provide both content and filter based recommender features, which mine information in the bibliographic records, holdings information, aggregated use data, and the like, to offer suggestions of other works of interest.

I.3   Support customization/personalization

I.3a: Allow user to define the set of resources/databases s/he wishes to search simultaneously, including a broader set of resources than those supported by current metasearch tools, such as Google restricted to .edu domains, museum and archive databases, and the like.

I.4   Offer alternative actions for failed or suspect searches

I.4a: Assess a user's input for likely spelling errors and offer alternatives, particularly if a term has few or no hits. Extend the services offered by general English-language systems such as Google to reflect the greater complexity of scholarly inquiry, including multi-lingual spell-checking and sensitivity to abstruse scholarly terms.

I.4b: Always offer constructive suggestions when a search produces zero results. Suggestions should include a broad range of options, including alternative search terms, related terms, options based on recommender features (ex: nothing on this topic found, would you be interested in this related topic?), offering to expand the search to other catalogs and/or WorldCat, offering to search Amazon or the Web, and options to get librarian assistance.

I.5   Offer better navigation of large sets of search results

I.5a: Implement FRBR concepts to present related works hierarchically, pulling together all records related to a particular work (e.g., Moby Dick), diverse expressions of that work (e.g., translations into German, Japanese and other languages), different versions of the same basic text (e.g., the Modern Library Classics vs. Penguin editions), and particular items (a copy of Moby Dick on the shelf).

I.5b: Follow all of the linking fields in serial records to present all of the variant titles to users in a "family tree."

I.5c: Implement faceted browsing based on sophisticated analysis of the contents of the records.

I.6   Deliver bibliographic services where the users are
I.6a: Enable library content and services to be integrated within campus virtual learning environments/course management systems (VLE/CMS), e.g., Sakai, WebCT, Blackboard, etc.

I.6b: Enable library content and services to be embedded in institutional portals.

I.6c: Expose our metadata to external search engines as thoughtfully as possible.

I.6d: Make our digital and unique collections available first within the UC community, then facing outwards.

I.7   Provide relevance ranking and leverage full-text
I.7a: Provide relevance ranking based on a broad set of criteria, to arrange a set of retrieved records so that those most likely to be relevant to the request are shown at the top of the retrieved set.

I.7b: Use full text for discovery and relevance ranking when available.

I.8   Provide better searching for non-Roman materials
I.8a: Provide better searching for non-Roman materials, allowing searching in both Roman and in the vernacular, sorting results in language-appropriate ways, and displaying results in both Roman and vernacular forms.

II.   Rearchitecting the OPAC
II.1   Create a single catalog interface for all of UC
II.1a: Create a single catalog interface for both local and system wide collections. Engage in a system wide planning process to identify the appropriate mechanism for implementing such a vision

II.2.   Support searching across the entire bibliographic information space
II.2a: Pre-harvest metadata for the entire bibliographic information space that represents UC library collections for ease of searching.

II.2b: Provide result sets arranged by format, grouped in terms of granularity and other facets, together with user options to rearrange the default order.

III. Adopting New Cataloging Practices

III.1 Rearchitect cataloging workflow

III.1a: View UC cataloging as a single enterprise, eliminating duplication and local variability in practice, agreeing on a single set of policies, sharing expertise, and maximizing efficiency. Engage in a system wide planning process to identify the appropriate mechanism for implementing such a vision.

III.1b: Implement a single data store for UC, be it a single file of cataloging records or the entire ILS.

III.2. Select the appropriate metadata scheme.

III.2a: Use level of description and schema (DC, LOM, VRA Core, etc,) appropriate to the bibliographic resource. Don't apply MARC, AACR2, and LCSH to everything.

III.2b: Consider the value of implementing the FAST syntax with special attention to 'place' and 'time periods' in order to support faceted browsing in those categories.

III.2c: Consider using controlled vocabularies only for name, uniform title, date, and place, and abandoning the use of controlled vocabularies [LCSH, MESH, etc] for topical subjects in bibliographic records. Consider whether automated enriched metadata such as TOC, indexes can become surrogates for subject headings and classification for retrieval.

III.2d: In allocating resources to descriptive and subject metadata creation, consider giving preference to those items that are completely undiscoverable without it, such as images, music, numeric databases, etc. Consider whether automated metadata creation techniques can be used for all textual materials.

III.3 Manually enrich metadata in important areas

III.3a: Enhance name, main title, series titles, and uniform titles for prolific authors in music, literature, and special collections.

III.3b: Implement structured serials holdings format.

III.4 Automate Metadata Creation

III.4a: Encourage the creation of metadata by vendors, and its ingestion into our catalog as early as possible in the process.

III.4b: Import enhanced metadata whenever, wherever it is available from vendors and other sources.

III.4c: Automate the addition of geographic data into our catalog to support existing services, and to support emerging services.

III.4d: Change the processing workflow from "Acquire-Catalog-Put on Shelf" to "Acquire-Put on Shelf with existing metadata-Begin ongoing metadata enhancement process through iterative automated query of metadata sources."

III.4e: Add enriched content such as Tables of Contents, cover art, publisher promotional blurbs, content excerpts (print, audio or video), and bibliographies. Build retrieval, relevance, and navigation services on top of this content.

IV: Supporting Continuous Improvement

IVa: Institutionalize an ongoing process of identifying and prioritizing improvements to our bibliographic services, in such a way that we get more than incremental improvements. Must lead to action, not just study. One task might be to track environmental scans, for example.

IVb: Provide robust reporting capability (data warehouse).

Introduction

Society is in the midst of learning how to "be" in the information age. The advent of computers and the inclusion of the Web in our work and private lives have pushed innovations and embraced information and access in ways we can hardly imagine. We are living in a complex and challenging digital landscape that changes constantly.

On the Library front, our bibliographic systems have not kept pace with this changing environment. The continuing proliferation of formats, tools, services, and technologies has upended how we arrange, retrieve, and present our holdings. Our users expect simplicity and immediate reward and [Amazon](), [Google](), and [iTunes]() are the standards against which we are judged. Our current systems pale beside them.

The current Library catalog is poorly designed for the tasks of finding, discovering, and selecting the growing set of resources available in our libraries. It is best at locating and obtaining a known item. For librarians and for our users, the catalog is only one option for accessing our collections. We offer a fragmented set of systems to search for published information (catalogs, A&I databases, full text journal sites, institutional repositories, etc) each with very different tools for identifying and obtaining materials. For the user, these distinctions are arbitrary.

Within Library workflows and systems too much effort is going into maintaining and integrating a fragmented infrastructure. We need to look seriously at opportunities to centralize and/or better coordinate services and data, while maintaining appropriate local control, as a way of reducing effort and complexity and of redirecting resources to focus on improving the user experience.

Books are not going away. Traditional information formats are, however, being used in combination with a multitude of new and evolving formats. It is our responsibility to assist our users in finding what they need without demanding that they acquire specialized knowledge or select among an array of "silo" systems whose distinctions seem arbitrary.

What the Future Holds

The famous sage Howard Cosell once said, "What's popular isn't always right. What's right isn't always popular." We suspect when it comes to the Internet and how it has simplified searching, what is popular is also right. Below we examine what users expect from the next generation library search interface and what infrastructure changes libraries need in order to continue to provide effective services. None of the bibliographic services envisioned below are Buck Rogers-like fantasies. Rather, the examples are found repeatedly throughout library literature, user surveys, and currently available technologies.

The examples below bring home what we need to offer if we expect to attract students and researchers into our collections and recruit innovative librarians into the field.

What Users Want

Users want a rich pool from which to search, simplicity, and satisfaction. One does not have to take a 50-minute instruction session to order from Amazon. Why should libraries continue to be so difficult for our users to master?

Examples:

- Users expect one system or search to cover a wide information universe (ala Google or Amazon.com).
- Enriched metadata. (ONIX, tables of contents, cover art, etc.).
- Full-text availability.
- Users want to move easily/seamlessly from a citation ABOUT an item to the item itself. Discovery alone is not enough.
- Users expect systems to provide lots of intelligent assistance.
  - Correct obvious spelling errors.
  - Sort results in order of relevance to their queries.
  - Help in navigating large retrievals through logical subsetting or topical maps or hierarchies.
  - Help in selecting the best thing through relevance ranking or added commentary from peers & experts or "others who used this also used that" tools.
  - Customization and personalization services.
- Authenticated single sign-on.
- Security/privacy.
- Communication and collaboration.
- Multiple formats available: e-books, mpeg, jpeg, rss and other push technologies – along with traditional, tangible formats.
- Direct links to E-mail, Instant Messaging (IM), sharing.
- Scholars increasingly participate in online virtual communities for research and education.
- Users want what the library has to offer, without having to come to the library to get it.

What Libraries Need

In order to offer the UC community the best possible service and access to the highest quality information, the UC Libraries need to look closely at how we manage entry points into our collections, catalogs, and finding aids.

The time and energy required to do Library business is unsustainable. We have people performing duplicative work throughout our system. We are unable to share matching resources or records across our multiple catalogs, content management systems, and differing standards. These redundancies have opportunity costs in terms of services we do not have the time or staff to offer. We all agree that the cost of our Bibliographic Services enterprise is unsupportable as we move into an increasingly digital world, yet a solution is nowhere in sight.

If we wish to remain a contender in the information marketplace, we need to incorporate efficient ways for obtaining, creating, and exporting metadata. We must respond to demands to enrich our data in new ways, to add value and provide unique services to our users, without draining our budget. Given its prohibitive cost, staff created metadata should be applied only when there is proven value for current and future scholars.

In addition to staff created resource descriptions, metadata can be obtained from vendors and publishers, derived automatically from data, or contributed by users. Whatever metadata exists in bibliographic systems needs to be fully utilized. Libraries need to actively exploit existing metadata and develop other information-rich conduits, which will enhance the number of searchable access points. Through this model of deep indexing we can revitalize our services and offer unique, efficient, and necessary services to our users.

Our challenge is to prioritize what work we can continue to do and then to do it intelligently, do it once throughout the entire system, and do it as well if not better than our "for profit" peers.

Examples:
- One virtual system
  - Build efficiencies into the catalog.
  - Be able to drill-down the bibliographic hierarchy
  - Build the ability to share information in and out of our systems.
- Develop collaborative tools and data for analysis.
  - Ability to determine detailed holdings on each campus.
  - Collection management/acquisitions.
  - Electronic resources management.
  - Shared print and shared print archives.
- Integration of bibliographic universe
  - Integration with course reserves
  - Integration with repositories
  - Tools to manage data for local consumption
- Accommodate non-MARC metadata schema and structures

- Capitalize on already-existing metadata, passing along enrichments like tables of contents, cover art
- Collaborate with other cultural heritage organizations to make accessible additional resources
- Push content out to the settings where users search for information.
- Simplify the process for finding information where it resides.

Only through knowing our audience, respecting their needs, and imaginatively reengineering our operations, can we revitalize the library's suite of bibliographic services.

Recommendations

"Things should be made as simple as possible, but no simpler."   *– Albert Einstein.*

T he University of California Library system seeks to maintain and enhance the highest quality academic research experience possible. Beginning in 1868, the State of California has been deeply dedicated to building a world-class library. There is no denying that the way people use libraries has changed as much as the purpose of libraries have stayed the same. Below are listed the Bibliographic Services Task Force's core recommendations for actions we must undertake if we are to remain viable in the information marketplace. Additional ideas for improving bibliographic services are included in Appendix F.

I. Enhancing Search and Retrieval

Library bibliographic services are a core element of the research endeavor. We provide tools in the form of catalogs, indexes & abstracts, and web pages that link our users to the materials they want. What we fail to provide is seamlessness, simplicity, and common language searching. For the past 10 years online searching has become simpler and more effective everywhere, except in library catalogs. Below are enhancements we can incorporate.

I.1 Provide users with direct access to item

According to the BSTF design principles, the goal of any successful search is to have few (three or less) clicks employed in obtaining a call number and/or full-text information. Users want immediate satisfaction. We know from our experiences that undergrads want "it" now, no matter what. If they can't get the authoritative full-text they'll take the second or third-best full-text resource. Faculty and grad students are currently willing to wait because they understand they must, but they too value faster fulfillment. If we don't do a better job of meeting increasing expectations, we will lose our users.

When full-text is available, a user expects a single click to bring up the full text. Our users deserve something better than the current system of confusing menu choices, links which

connect to a journal site (and not a specific article), and URLs which are not stable and can't be bookmarked.

For print materials, our systems should offer a wide range of options when immediate delivery is not possible, including information about how long each option will take for fulfillment. It is up to us to design a system that looks for alternatives on our users' behalf. We need to move beyond Request and Recall as our only options for fulfillment of print materials, especially when the turn around time will not be fast enough to meet the user's need. For example, when a book is checked out, the system could identify a copy in a library across town via OpenWorldCat, identify a copy for purchase in the campus bookstore, initiate an immediate order via on online bookseller, and offer similar titles that are available on the shelf. When an item is on the shelf, the system could show a map of the location, offer advice on which locations are closest, or offer delivery services for a fee. Users should have enough information to make an informed choice of what option to follow, including those that take longer and those that involve a cost.

I.1a: RECOMMENDATION: Have UC eLinks take you to a logical, default choice, with option to go back to the menu if you want a different option. (If there is a reliable full-text link that would be first choice. This assumes that in the majority of times, we could correctly anticipate what service the user would want.)

I.1b: RECOMMENDATION: Provide an "I-want-this" button that is present when the context warrants, with the goal of always offering a fulfillment option. No dead ends. Give the user an option to specify turnaround time; work behind the scenes to fulfill as well as we can.

I.2 Provide recommender features

Our users increasingly expect online systems to provide more than a literal response to a query. Bibliographic systems must add value to the interaction, using what is known about both the user and his/her request to provide intelligent advice and assistance. Many Internet services make recommendations based on search history and an analysis of records retrieved and viewed. UC bibliographic systems have access to a wealth of data on which we could base recommendations with scholarly depth and significance. Though still experimental and preliminary, the Mellon-funded study being conducted by CDL has shown some promising results.

Recommender features offer the possibility of recommending to users a targeted set of alternative results ("recommendations") that are deemed likely to meet the users' need for information. Such recommendations can be content-based, from an analysis of content of a user's retrieved set, or filter-based, from an analysis of what other similar people have chosen as relevant or interesting. Works recommended could include:
- Items classified together and in the same building
- Items classified together regardless of physical location
- Items by same author
- Items on same subject

- Items about the same (or a nearby) place
- Items from the same (or a nearby) place
- Other editions of same work
- Other versions of a journal article (pre- and postprint)
- Other works which were checked out or viewed by other people in the same category of user (faculty, graduate student, etc) or by people in the same discipline/department.

I.2a: RECOMMENDATION: Provide both content and filter based recommender features, which mine information in the bibliographic records, holdings information, aggregated use data, and the like, to offer suggestions of other works of interest.

Examples: Amazon - http://www.amazon.com

I.3 Support customization/personalization

Bibliographic systems should allow users to assert some control over their research interactions. At minimum, a user should be able to set up a permanent profile of preferences that affects future uses of the system. Such a feature is particularly valuable as our systems grow and expand into more of the information space.

I.3a: RECOMMENDATION: Allow user to define the set of resources/databases s/he wishes to search simultaneously, including a broader set of resources than those supported by current metasearch tools, such as Google restricted to .edu domains, museum and archive databases, and the like.

I.4 Offer alternative actions for failed or suspect searches

Analysis of search logs shows that many users make predictable and correctable mistakes in searching. Intelligent systems will recognize such likely errors at input and offer suggestions to correct. User-focused systems will never leave a user alone with failure, facing zero results with no alternative path.

I.4a: RECOMMENDATION: Assess a user's input for likely spelling errors and offer alternatives, particularly if a term has few or no hits.   Extend the services offered by general English-language systems such as Google to reflect the greater complexity of scholarly inquiry, including multi-lingual spell-checking and sensitivity to abstruse scholarly terms.

I.4b: RECOMMENDATION: Always offer constructive suggestions when a search produces zero results. Suggestions should include a broad range of options, including alternative search terms, related terms, options based on recommender features (ex: nothing on this topic found, would you be interested in this related topic?), offering to expand the

search to other catalogs and/or WorldCat, offering to search Amazon or the Web, and options to get librarian assistance.

I.5   Offer better navigation of large sets of search results

Unlike casual searchers, scholarly researchers often need to work through a large set of search results. Relevance ranking is not enough when hundreds of records are germane. Large results sets should be presented in logical subsets, preferably through a dynamic process that responds to user choices with ever more useful drill-down options.

Navigating large sets of search results is particularly frustrating when many of the records retrieved are variations of the same work. When each edition of Hamlet, each translation, and each filmed version of the play is listed as a separate record, a user can be faced with screen after screen of displays of titles, each of which a user has to investigate to determine if the record is really what s/he needs. IFLA's Functional Requirements for Bibliographic Records (FRBR) defines a set of relationships that can be used to organize the records more logically. FRBR concepts differentiate records related to a particular work, diverse expressions of the work, different versions of the same basic text, and particular physical items.

Similarly, when a serial has had many title changes, it is very hard to construct a search in today's catalogs which will retrieve all of the records for all the title variants. Even if all the records are retrieved, the user must look at each record and traverse the arcane landscape of "Continues" and "Continued By" fields to understand the relationships among the titles.

Another method for helping users to navigate large sets of search results is to break the set into logical subsets, or "facets." In faceted browsing, the retrieved set is analyzed and organized based on the most appropriate information in the records. Some of the subsets most commonly used for faceting include
- Date
- Language
- Format
- Subject headings
- Name headings
- Availability (whether checked out, missing, on the shelf, etc)

When appropriate, subsets should be further subdivided into ranges (e.g., dates by decade, geographic locations by region, etc.). The most sophisticated systems analyze the retrieved set semantically and subset by how many records are under a particular topic. For instance, if hundreds of records in the set fall into the date range of 1990-2000, the system would subset by individual year.

I.5a: RECOMMENDATION: Implement FRBR concepts to present related works hierarchically, pulling together all records related to a particular work (e.g., *Moby Dick*), diverse expressions of that work (e.g., translations into German, Japanese and other

languages), different versions of the same basic text (e.g., the Modern Library Classics vs. Penguin editions), and particular items (a copy of *Moby Dick* on the shelf).

I.5b: RECOMMENDATION: Follow all of the linking fields in serial records to present all of the variant titles to users in a "family tree."

I.5c: RECOMMENDATION: Implement faceted browsing based on sophisticated analysis of the contents of the records.

Examples of FRBRization:
        OCLC's Fiction Finder Project
        http://fictionfinder.oclc.org/

        RLG's RedLightGreen
        http://www.redlightgreen.com

Examples of faceted browsing:
        Endeca's ProFind
        http://endeca.com/demos/index.html

        Elsevier's Scopus
        http://www.info.scopus.com/demo/

        North Carolina State University new catalog
        http://www.lib.ncsu.edu/catalog/

        RLG's RedLightGreen
        http://www.redlightgreen.com

Example of serial record "family tree":
        http://www.secstate.wa.gov/library/docs/iii/seattlepi.htm

Relevant references:
Gonzales, Linda (Apr 15, 2005) What is FRBR?.

Yee, Martha M. (2005) FRBRization: a method for turning online public finding lists into online public catalogs.

I.6   Deliver bibliographic services where the users are

Library content and services must be made available outside of library systems and websites allowing for users to access them more conveniently at the point of need --within their preferred work and/or study environment. We should strive to make information available at the point the user requires it. This vision contrasts with the more traditional approach of building dedicated Web sites and expecting the user to find and access resources within them. Instead the library experience should be reproduced wherever and whenever the user

requires it, without the need to visit a separate website for the library. This way of thinking actively encourages the use and re-use of library resources.

Many campuses, for example, are making use of virtual learning environments/course management systems (VLE/CMS). These campuses are also starting to implement institutional portals to facilitate the aggregation and presentation of campus applications and services to their staff and students.   Instead of requiring our users to access library resources through our own website we can present relevant information, directly within these environments. At minimum, we can include tools such as the library search box, recommended resources, news, and notification services within the systems.

The technical framework that enables us to create this new environment, which Lorcan Dempsey refers to as the 'recombinant library', includes open standards, such as XML, RSS, WSRP, and JSR168, and new concepts which are increasingly being dubbed as Web 2.0. Through Web Services and other protocol-based integration, library collections and services can be incorporated transparently, appearing as features within other systems.

Looking even further afield, library content and services need not remain hidden on our own campuses, available only within our own interfaces. As our users move outside the library catalog's walled gardens, it behooves us to meet and support them, no matter where they discover our resources. We recognize by now that our users are frequently using search engines to find information. We should ensure that our metadata is exposed to all search engines that want it.   Given the widespread duplication among research libraries, having each library expose MARC data individually raises the possibility of diluting search results. Rather, we recommend working with OCLC to apply its OpenWorldCat technique to expose all of its records, or at least all UC records, for discovery by search engines. Searchers can retrieve "find in a library" results via searches that include or are limited to, "worldcatlibraries.org". It's also important to expose our metadata so that we can support additional services we may want to provide in the future.

In addition to our MARC records, we should also expose the metadata from our digital collections. Each campus has unique materials that are valuable and would benefit the UC community at large if they were discoverable. There are several mechanisms for doing so, including OAI and Google, each with its own promise, and each with its own set of challenges.

I.6a: RECOMMENDATION: Enable library content and services to be integrated within campus virtual learning environments/course management systems (VLE/CMS), e.g., Sakai, WebCT, Blackboard, etc.

I.6b: RECOMMENDATION: Enable library content and services to be embedded in institutional portals.

I.6c: RECOMMENDATION: Expose our metadata to external search engines as thoughtfully as possible.

I.6d: RECOMMENDATION: Make our digital and unique collections available first within the UC community, then facing outwards.

Relevant references:
Awre, Chris et al. (Oct 2005) The CREE Project: investigating user requirements for searching within institutional environments.

Dempsey, Lorcan. (2003) "The recombinant library: portals and people."

I.7   Provide relevance ranking and leverage full-text

Currently our users type terms into a search box and if they are lucky they retrieve a set of records that are close to the terms they typed in. The good news is that in some systems the most current titles appear first in the results set. The bad news is that currency is often not the best determinant of relevancy or success.

Relevance ranking within most library catalogs is keyword or LCSH based, but we could learn many lessons from Google/Yahoo/Amazon-like search engines.

We have options as to how we can weight titles with criteria that will enrich the search results:
> How often has an item been on class reserves lists
> Circulation frequency
> Citation analysis
> Number of institutions holding the item

We could extend our ability to search and retrieve into the full-text content of a book. Using this content intelligently we could choose significant, information-rich sections such as TOC, abstracts, notes, and bibliographies to weight retrieval and to guide relevance ranking.

I.7a: RECOMMENDATION: Provide relevance ranking based on a broad set of criteria, to arrange a set of retrieved records so that those most likely to be relevant to the request are shown at the top of the retrieved set.

I.7b: RECOMMENDATION: Use full text for discovery and relevance ranking when available.

I.8   Provide better searching for non-Roman materials

We are a library system with a strong non-English language collection. Our research libraries, so diverse and with so many languages, have to do a better job providing access to our foreign language collections. We need to improve our support of users and scholars who research in the vernacular.

I.8a: RECOMMENDATION: Provide better searching for non-Roman materials, allowing searching in both Roman and in the vernacular, sorting results in language-appropriate ways, and displaying results in both Roman and vernacular forms.

II. Rearchitecting the OPAC

In today's climate we need to radically rethink what is in our catalog, how it is searched, and connectivity to individual items.

II.1 Create a single catalog interface for all of UC

Given the increase in shared collection building, we should be presenting a single catalog interface for both local and system wide collections. In addition to being an improvement in user service, a single catalog yields significant efficiencies and cost savings for libraries. "Doing more with less" has been a reality for the University of California library systems over the past decade. There is huge overhead in maintaining 10 campus OPACs, and overhead in the extraction, transmission, ingestion, merging and indexing of records to provide the union catalog experience. A single OPAC would both provide a single point of entry for users into UC's rich resources and produce efficiencies and cost savings.

To gain the maximum efficiency in creating and supporting a single UC wide OPAC, there should be a parallel effort to create and support a UC-managed catalog database (see III.1 below). By creating a single data store (the database) for bibliographic records, we can more easily create discovery and presentation services that lay on top of the data store. Indeed, one can imagine a researcher searching her catalog from a campus, and having the results presented to her at different levels of physical location of the materials (campus, UC-wide, ILL).

The technical framework that enables us to create a single UC OPAC and database could take several different forms. UC could decide to standardize on a vendor catalog product hosted within UC, and create the supporting policy framework to make this successful. Or, UC could decide to make use of an outside vendor to host the database, and/or create the union view. For example, UC could contract with RLG or OCLC to host our database and to create a specialized union view based on our specifications. OCLC has been prototyping services that UC would be interested in seeing as production level services (FRBR, enriched metadata such as cover art, user-provided reviews) that UC could benefit from. Or, OCLC could host the database, and UC could provide the union view.

While there are great efficiencies to be gained in supporting one catalog, and one OPAC, we recognize there are concerns with this approach. Which vendor's cataloging product would we choose, and where would it be hosted? How could we centralize operations and personnel, given that individual campuses have budgetary constraints and practices that limit such practices? Given OCLC's ever-increasing costs, would it make sense to work with them as a partner? If we have only a union catalog, what about campus-specific presentations of the catalog?   More debate and discussion is needed to identify the best option for presenting a single point of entry for our users. Some of the options we explored include:

- Single UC OPAC plus other resources.
    - Pros
        - Significant savings with one system
        - Seamless interface for users doing cross-campus research
        - Similarity of system allows for ease of shared projects
    - Cons
        - Systems weaknesses are similar across UC
        - If we use a vended system and the company fails, so goes our support
        - No one system fits our divergent desires

- Outsource the UC OPAC (OCLC, RedLightGreen, Google, etc)
    - Pros
        - Places the maintenance of system off campus
        - Allows for simplified cataloging processes
        - Allows us to take advantage of their innovation, eg OCLC's OpenWorldCat, user provided reviews, etc.
    - Cons
        - Unproven ability to support holding records
        - Unproven ability to support user-initiated services

II.1a: RECOMMENDATION: Create a single catalog interface for both local and system wide collections. Engage in a system wide planning process to identify the appropriate mechanism for implementing such a vision

Relevant references:
Byrum, John D Jr. (2005). Recommendations for urgently needed improvement of OPAC and the role of the National Bibliographic Agency in achieving it.

Pace, Andrew K. (Feb 2005) My kingdom for an OPAC

II.2   Support searching across the entire bibliographic information space

Users are often unaware that there are multiple discovery tools for the resources the library has to offer: the library catalog, abstracting and indexing databases, the e-Scholarship Repository, various collections of digital library objects, archival collections, etc. As a result, they are frequently frustrated by their lack of success in finding what they seek. The few sophisticated researchers who are aware of the differences are justifiably unhappy with the need to search one "silo" at a time.

Users who are accustomed to Google expect to enter one search and retrieve information pulled together from across the information space and presented in a single ranked list. They want more than the ability to search multiple catalogs or multiple A&I databases simultaneously. They expect to search the full range of tools cited above or subsets the user wishes to select.

A broadcast metasearch tool has built-in limitations in providing a true federated search across such a wide set of resources. Such a tool is at the mercy of the varying capabilities and reliability of the distributed systems it targets, and the services it provides are tied to the lowest common level of protocol and searching those systems provide. In order to support a predictable federated search across a wide set of resources, and to be able to build high-level searching and display services based on that searching, a federated search tool should pre-harvest as much metadata as possible. Google has shown the power of pre-harvesting metadata, applying sophisticated processing to the metadata, and building a coherent set of services on top. We should create a similar set of services by pre-harvesting metadata for the full set of UC Library collections.

Within such a pre-harvested federated search service, search results must be presented to users in a logical way. A default display could present faceted results by type of material (e.g., journals, monographs, images, sound files, etc.).

For example, a search for a world leader might return
- books and entire journals about the leader
- articles about the leader
- conference papers and presentations
- photographs
- editorial cartoons
- digitized letters the leader wrote or received

It would then be possible for a user to re-sort the results according to other facets, such as language, date, or place of issuance.

II.2a: RECOMMENDATION: Pre-harvest metadata for the entire bibliographic information space that represents UC library collections for ease of searching.

II.2b: RECOMMENDATION: Provide result sets arranged by format, grouped in terms of granularity and other facets, together with user options to rearrange the default order.

III. Adopting New Cataloging Practices

Traditional library cataloging has always been relatively expensive to create and time-consuming to maintain. Being that the majority of the traditional cataloging enterprise dealt with commodity material (identical items owned by many libraries), shared copy cataloging systems such as OCLC and RLG offered efficiencies and kept the effort manageable.

However, as huge amounts of e-learning items and unique digital materials are added to our collections, the sheer volume, diversity, and complexity of such materials will require new forms of cataloging practices to be adopted.

III.1 Rearchitect cataloging workflow

Cataloging and metadata expertise represents a scarce and valuable resource which the University of California must continue to maximize. A report from the OCLC collection analysis tool recently found that 77% of the monographs published in the last 10 years are held on more than one campus, suggesting that a significant amount of duplicate acquisitions and cataloging is taking place within the system. At the same time, many parts of our collection suffer from inadequate bibliographic control, represented by backlogs in foreign language cataloging, minimally cataloged records sent to the RLFs, and hidden treasures only found within digital collections.

To maximize the effectiveness of our metadata creation, University of California cataloging should be viewed as a single enterprise. We need to move beyond a practice of shared cataloging to a practice of integrated cataloging, in which the system adopts a single set of cataloging standards and policy, eliminates duplication of effort and local variability in practice, provides system wide access to language, format, and subject expertise, and creates a single copy of each bibliographic record for the entire system.

There are several options for implementing such a single enterprise vision and the task force did not agree on a single option to recommend. One could create the single enterprise virtually, physically, or as a combination of the two. More debate and discussion is needed to identify the appropriate mechanism. We explored the following options, though others are also feasible.

Organization options:

- Coordinate cataloging expertise and practice across the entire system. Expand the model of the Shared Cataloging Program, whereby material is cataloged one time for all parties with access to the material, to include difficult-to-find format, language, and subject expertise.

- Consolidate cataloging into one or two centers within UC. By pooling cataloging experts in one place, a wider proportion of our collections could be cataloged quickly, utilizing shared expertise. Some of the logistical challenge of getting material to the catalogers could be addressed if the centralization of acquisitions were coordinated with that of cataloging. In a single unit, it would be easier to create standardized policies, review output for consistency, and adapt to change in a unified way. On the other hand, if one were able to eliminate duplication of effort and local variation within the UC System that might yield equivalent efficiencies without consolidation. Reducing cataloging and metadata expertise on individual campuses could hamper our ability to respond to new local initiatives, tap metadata expertise for campus projects, and assist public services staff to use bibliographic systems effectively.

- Outsource a greater proportion of standard cataloging work. UC libraries already use outsourcing to handle languages and formats in which the library lacks expertise. If we outsource the majority of standard MARC cataloging, library staff could focus on new and expanded uses of metadata. The cost of such outsourcing, though, might outweigh the benefits.

Architecture options:

- Create a shared central file with a single copy of each bibliographic record, and attach all holdings in the UC system. Adoption of a single file of catalog records would eliminate the need for algorithms to merge copies of records on the fly in the union catalog. A data flow from the central file to the campus ILSs would mean that any bibliographic record update or enrichment would automatically be shared with all campuses. This option changes the direction of data flow from the central system to local systems, increasing the need for local loading and merging.

- Adopt a single ILS for the entire University of California system. Storing all records in a single system would eliminate the need for duplicate data, complex data flows, and complicated merging algorithms. A single system could also improve support for cooperative collection development activity. Making decisions together and adhering to system-wide standards improves the services we can offer our users. Supporting local inventory control and user-initiated services could be more complex in a single-system model.

- Rely on OCLC as the single UC database of record for bibliographic data. This option is appealing if it could include automatic access to the latest copies of records. The ability to edit records would be limited to those levels of staff authorized to replace master records in the bibliographic utility. Implementing any UC-only data standards or recording copy-specific notes would be difficult.

III.1a: RECOMMENDATION: View UC cataloging as a single enterprise, eliminating duplication and local variability in practice, agreeing on a single set of policies, sharing expertise, and maximizing efficiency. Engage in a system wide planning process to identify the appropriate mechanism for implementing such a vision.

III.1b: RECOMMENDATION: Implement a single data store for UC, be it a single file of cataloging records or the entire ILS.

Examples of a single data store:
Österreichischer Bibliothekenverbund, The Austrian Library Network
a consortium of academic libraries using a single file of cataloging records
http://www.obvsg.at/

University System of Maryland and Affiliated Institutions (USMAI)
a consortium of libraries using one ILS, implemented as a single file
http://www.itd.umd.edu/

III.2. Select the appropriate metadata scheme.

Such alternate schemes as Dublin Core, Learning Object Metadata (LOM), and Visual Resources Association (VRA) Core standards offer simpler ways of describing the range of new resources in our collections than traditional cataloging. They propose a core set of metadata elements that emphasize retrieval, rather than description, an effort that is obviously not as important in a world where the item is made directly accessible to users on a computer terminal.

Another response to the need for simpler and more efficient cataloging practices is OCLC's Faceted Application of Subject Terminology (FAST) based on the Library of Congress Subject Headings schema (LCSH). LCSH's complex syntax and rules for constructing headings restrict its application by requiring highly skilled personnel and limit the effectiveness of automated authority control. To make bibliographic control systems easier to use, understand, and apply, OCLC has modified the LCSH with a simpler syntax. FAST retains the very rich vocabulary of LCSH while making the schema easier to understand, control, apply, and use.

Dublin Core, LOM, VRA Core, and FAST are much easier to use than traditional cataloging techniques based on MARC , AACR2, and LCSH. Rather than focusing on description which only humans can interpret, metadata in these schemes is segmented in ways that allow systems to act upon them directly, enabling enhanced retrieval, new tools for browsing, and Web-based access to records and services.

The task force agrees that our bibliographic systems should accommodate multiple metadata schemes. We also agree that controlled vocabularies are still very valuable for name, uniform title, date, and place. Not all agree, though, that the current controlled vocabularies are as effective for topical subjects. The different points of view which arose during our discussions included:

- As we import or link to more full text and enhanced descriptive metadata, apply sophisticated algorithms to that metadata, and provide richer retrieval and browsing options, using controlled vocabularies such as LCSH and MeSH for topical subjects is no longer as necessary or valuable. Given our limited cataloging resources, we should apply subject analysis only to material that is not self-discoverable through textual searching. Where controlled vocabulary is used, we should replace the traditional LCSH structure with a more structured syntax such as FAST, which is more machine-actionable.

- Even with full text searching and enhanced metadata, topical subject headings still provide a valuable collocation service when searching large collections, particularly in multiple languages. Though machine algorithms might deal successfully with synonyms and related terms in one language, they are less likely to be as successful across many languages. Also, deviating from standard cataloging practice could limit

our ability to both share our cataloging records and import cataloging records from others.

More debate and discussion is needed to identify the appropriate strategy for description, subject access, and co-location in a full-text world.

III.2a: RECOMMENDATION: Use level of description and schema (DC, LOM, VRA Core, etc,) appropriate to the bibliographic resource. Don't apply MARC, AACR2, and LCSH to everything.

III.2b: RECOMMENDATION: Consider the value of implementing the FAST syntax with special attention to 'place' and 'time periods' in order to support faceted browsing in those categories.

III.2c: RECOMMENDATION: Consider using controlled vocabularies only for name, uniform title, date, and place, and abandoning the use of controlled vocabularies [LCSH, MESH, etc] for topical subjects in bibliographic records. Consider whether automated enriched metadata such as TOC, indexes can become surrogates for subject headings and classification for retrieval.

III.2d: RECOMMENDATION: In allocating resources to descriptive and subject metadata creation, consider giving preference to those items that are completely undiscoverable without it, such as images, music, numeric databases, etc. Consider whether automated metadata creation techniques can be used for all textual materials.

Relevant references:
Hyatt, Shirley. (2003) "Developments in Cataloging and Metadata."

Lynch, C. (Nov 2000) The New context for bibliographic control in the new millennium.

Mann, T. (Aug 2005) Will Google's keyword searching eliminate the need for LC cataloging and classification?

III.3   Manually enrich metadata in important areas

Manual metadata creation is by definition both expensive and time-consuming and is an activity that should judiciously be applied where it yields the most benefit. There are a number of areas where the application of intellectual effort in the creation of metadata justifies the high cost.

The enhancement of FRBR relationships through the manual addition or correction of name, main title, series titles, and uniform title, especially for prolific authors in the fields of music, literature, and special collections is one such area. The collocation of materials and the concomitant search and retrieval improvements in these fields more than justify the cost.

Additional attention to serials holdings would likewise have a major positive impact of effective search and retrieval. If serials holding were better structured, services to users would be much more reliable and major efficiencies could be reached through automated record matching and processing.

III.3a: RECOMMENDATION: Enhance name, main title, series titles, and uniform titles for prolific authors in music, literature, and special collections.

III.3b: RECOMMENDATION: Implement structured serials holdings format.

III.4 Automate Metadata Creation

As increasingly more data becomes available to the UC community, we need to take advantage of automated tools and processes that enable us to create metadata, or help speed the metadata into our catalogs.

While we already import and load MARC records supplied by vendors at the point of acquisition (e.g. PromptCat), we should encourage more vendor participation. On our end, we should change our practices to accept the records as provided, with no enhancements. We must adapt and recognize that "good enough is good enough", we can no longer invest in "perfect" bibliographic records for all materials. It can be helpful to think of metadata provision as an ongoing process versus a one-time event. Materials can be provided to users shortly after receipt, with whatever metadata is available at the time, with the assumption that metadata would be successively updated and upgraded over time as automated metadata becomes available. We should accept skeletal records when available (e.g., title lists from content aggregators), and enhance skeletal, or minimal, records through iterative automated queries of metadata content. If the material is high-value enough to justify it, additional human intervention is also an option.

Incorporating metadata into the process earlier rather than later results in increased access, even with imperfect metadata. For example, Lexis/Nexis is an aggregation of selected full-text content from many sources such as individual newspapers and journals. If we scripted a process to load skeletal bibliographic records for each of the Lexis/Nexis sources into the catalog, our users would have some chance of discovering that content in a catalog search, however skeletal the records. Such skeletal records would also help explain perceived gaps in serials holdings. On the other hand, if there is minimal metadata, this would have a deleterious effect on record merging, and the implementation of FRBR.

One of the reasons that Amazon is popular and successful is that Amazon includes enriched metadata, such as Tables of Contents, cover art, publisher promotional blurbs, content excerpts (print, audio or video), and bibliographies. Other libraries are adding much of this content to their bibliographic systems. The University of California should follow suit. This content can enhance retrieval, relevance ranking, and recommender features. According to some studies, faculty and graduate students use bibliographies as primary research tools and entry points into a field.

Finally, we recommend mapping geographic terms to latitude and longitude, using GNIS, if it can be done in an automated way. The addition of geographic data can support retrieval, GIS-based search interfaces (e.g., give me items close to where I am), support integration with Google and other mapping services, and support a hierarchical approach to place names (drill down to more specific locations, or expand to broader areas).

III.4a: RECOMMENDATION: Encourage the creation of metadata by vendors, and its ingestion into our catalog as early as possible in the process.

III.4b: RECOMMENDATION: Import enhanced metadata whenever, wherever it is available from vendors and other sources.

III.4c: RECOMMENDATION: Automate the addition of geographic data into our catalog to support existing services, and to support emerging services.

III.4d: RECOMMENDATION: Change the processing workflow from "Acquire-Catalog-Put on Shelf" to "Acquire-Put on Shelf with existing metadata-Begin ongoing metadata enhancement process through iterative automated query of metadata sources."

III.4e: RECOMMENDATION: Add enriched content such as Tables of Contents, cover art, publisher promotional blurbs, content excerpts (print, audio or video), and bibliographies. Build retrieval, relevance, and navigation services on top of this content.

IV: Supporting Continuous Improvement

Technology revolutionizes itself every 6 months. We libraries can no longer afford to sit back and enjoy our position as supreme bibliographers and catalogers. The private sector is beating us at our own game through their focus on research & development. We do not have to do our own environmental scans and reinvent the wheel; instead we can track national trends, read and evaluate, and act. If we don't continue to track and set priorities for change after the work of the BSTF ends, we will stagnate again. Don't let the momentum end.

Many Integrated Library Systems do not provide the deep level of inventory reporting many selectors desire. It would be extremely powerful if we could pull reports from the system that are meaningful to the selectors. Using the catalog to build reports on what books have been searched, read online, checked out, considered but ignored, etc. These are all bits of information that can help a selector build a stronger, more heavily used and relevant collection.

IVa: RECOMMENDATION: Institutionalize an ongoing process of identifying and prioritizing improvements to our bibliographic services, in such a way that we get more than incremental improvements. Must lead to action, not just study. One task might be to track environmental scans, for example.

IVb: RECOMMENDATION: Provide robust reporting capability (data warehouse).

Scenarios

W hat will it take to move the UC Libraries into the future? How do we stay true to our strengths while incorporating the best of technology and the do-it-yourself attitude of today's Web-empowered researchers?

While most people understand the benefits of developing a vision of the future, achieving the right balance between what is technically possible and what is desirable can be a difficult task. Below are three scenarios that share the same non-negotiable principles. The scenarios represent matching standards or points of action for which there is a modest, moderate, and radical approach.

Scenario planning exercises are used in many large organizations to build understanding of what the impact of different possible futures might be. Within an organization, scenarios provide a common vocabulary and an effective basis for communicating complex conditions and options. The result is a small set of internally consistent, but substantively different, scenarios that can be considered alongside each other.

There is no reason to limit oneself to the examples in one scenario, mash them up! For one of the examples one could favor a moderate approach and then choose a radical approach for the following example. The three approaches simply present the information in a shared context.

Please read the following, noting what sparks your excitement and what discomforts you.

Modest Scenario

User Experience	Library Changes to Implement the Scenario
Sarah, a UC Irvine junior, needs to begin research for a short paper on African political theory for her Political Science class. The assignment is to write a paper citing a combination of monographs and serials. She starts her research within the Library Catalog, one of the many library tools accessible via, Blackboard, the course management system (CMS) preferred by her professor.	- Deliver bibliographic services to where the users are (ex: integrate into VLE/CMS, portals, etc.). - Ensure all bib resources can be targeted via persistent URLs. - Become absolutely integrated within the structure of learning and teaching lest we lose our market share.
In the Catalog she discovers an array of books on her topic. They are found at a number of the UC libraries.	- View UC cataloging as a single enterprise. - Build a single cataloging system; create records once, attach holdings, and then export to local systems. - Eliminate the need to merge in Melvyl, thus speeding up and simplifying the process of upgrading to new catalog versions.
Excited about two titles in particular, Sarah is certain these books will provide the introduction to her topic she desires. By using key words and geographic fields she located a number of interesting titles. By clicking on the link to a title, an image of the book's cover appears adding another level of visual information. Sarah liked the subtitle and author blurb and decided to save the record to her CMS.	- Change how we do metadata – do less of some, more of others; automate more; enrich metadata (ex: with cover art, etc.). - Incorporate selective outsourcing; add publisher info, and free stuff such as imported vendor MARC records at time of acquisition. - Create local geographic fields from geographic parts of subject headings and import matching lat/long from GNIS.
Sarah especially appreciated the intelligent assistance. Originally she typed in "Zimbabway". The catalog corrected her spelling and offered up the proper name of the country (Zimbabwe)	- Improve our discovery tools. - Provide intelligent assistance such as spell check. - Always offer an alternative action for failed searches.
One of the first titles listed was over a decade old, but Sarah trusted the catalog had directed her to a list of titles that best fit her criteria. With the built in relevance ranking and recommender systems, the first title was more than likely the seminal text on Zimbabwe politics.	- Leverage metadata we have. - Ensure that we preserve usage data for recommender systems and relevance ranking while being sensitive to privacy concerns. - Ensure that all metadata licensed by any campus is available to all campuses. - No more piece meal improvements.

Sarah found the search results logical to peruse as they were divided into subcategories by subjects, authors, and languages	▪ Better navigation of results sets. ▪ Faceted browsing: analyze the retrieved set and present to searcher in logical subsets.
She clicked on one of the links and immediately, without any more clicks, opened up *A decade of democracy in Africa* [electronic resource] / edited by Stephen N. Ndegwa	▪ When possible provide direct access to item, with no intermediate menus. ▪ Include all digitized content (ex: Open Content Alliance) in bibliographic systems for direct access.
Within the Catalog search results there was also a good selection of articles from a selection of abstracts and indexes that would help supplement the monographic information.	▪ Support searching across the entire bibliographic information space: catalog, journal articles, digital collections, etc. ▪ Implement metasearch for the A&I databases (at the very minimum), allowing the user to define a set of resources/databases s/he wishes to search simultaneously.
Sarah finished her research and began the task of reading her books and articles, secure in the knowledge that The Library system is engaged in continual improvement	▪ System-wide commitment to continuous improvement. ▪ Plan on a periodic revision of this plan. ▪ Conduct periodic user assessments to inform the revised plan. ▪ Act on plans.

Moderate Scenario

User Experience	Library Changes to Implement the Scenario
Santos, an honors student in Folklore at UCB is working on his Senior Thesis: *The Persistence of Social Memory in Folklore*. He has never before coordinated such a complex project. To meet the requirements of a thesis he will need to incorporate primary and secondary resources. He is open to a variety of formats and will need to be deeply creative, being that his subject is cross-disciplinary and without any sort of controlled vocabulary.	
Beginning his background research, Santos turns to the UC single catalog where he is able to do a cleaner (more successfully merged files) system-wide search than previously possible with Melvyl. He quickly finds out that it is difficult to research this subject. The topic crosses into literature/folklore/sociology/ psychology and a number of other subjects. The search terms are inconsistent and a bit tricky. There are no LCSH heading that work well and the premise, while floating through many writings, is not an overt theme in most of the literature.	▪ View UC cataloging as a single enterprise. ▪ Build a single UC integrated library system with separate metadata stores for each campus, possibly hosted on a campus instead of at CDL. ▪ Careful governance planning, workflow design, links to campus systems, and support for local circulation. ▪ No data flow problems shipping bibliographic records to and from campus systems. ▪ Single acquisitions system facilitates collaborative collection development and acquisitions. ▪ Economies of scale and savings in staff time release campus metadata expertise to do more value-add services and development.

The new ILS provides the primary standard access points while expanding the number of searchable access points. Some of the new features are very slick. The ability to click and view book cover art and the inclusion of table of contents (TOC) and indexes for searching allows for natural language searching and a more successful hit ratio.	▪ Change how we do metadata – do less of some, more of others; automate more; enrich metadata (ex: with cover art, tables of contents, full text, etc). ▪ Cut back on controlled vocabularies for topical subjects; put extra effort into controlled vocabularies for name, uniform title, date, and place. ▪ Add geographic fields with lat/long to all applicable records. ▪ Support multiple metadata schemes (ex: MARC/DC, VRA, etc). ▪ Review minimum standards to ensure that we expose the records of our collections to the maximum extent possible, for maximum discovery of resources.
The improved discovery tools allow for a better result set, providing intelligent assistance.	▪ Ensure that we preserve usage data for recommender systems and relevance ranking while being sensitive to privacy concerns. ▪ Incorporate relevance ranking and recommender systems. ▪ Index TOC, abstracts, other enriched metadata for a wider variety of searchable metadata.

Once Santos retrieves his results set he is able to better navigate through the options.
    Ex: Social Memory (redlightgreen.org)
       Subjects:
          Memory - Social Aspects
          Memory
          Social Psychology
          Social Perception
          Electronic Books
       Authors:
          Halbwachs, Maurice
          Alexandre, Jeanne Halbwachs
          Bartlett, Frederic C Sir
          Bartlett, Frederic Charles
          Bartlett, Frederic Charles Sir
       Languages:
          English
          French
          German
          Chinese
          Portuguese
	▪ Leverage metadata we have.
	▪ Implement FRBR with metadata already in the records, to provide better-navigable groupings of related records for the user.
	▪ Implement faceted browsing, using metatdata already in the records to subset the retrieval in logical groupings.

Choosing books by Maurice Halbwachs, Santos is able to get direct access to the books held within the UC system. There is one book in particular he wants to read. It is checked out, but before Santos can despair the catalog directs him to the Stanford Library holdings where the title is not checked out. The other books he wants are easily requested from SRLF and UCD. The UCB holding record includes a link to a map of the stacks showing the exact location he can expect to find the books.	▪ Direct access to item. No dead ends; always offer a fulfillment option (ex: request, link to online booksellers, ask a librarian, find at another local library, etc.). ▪ Show maps in the library to direct user to the location on the shelf.
As far as Santos is concerned, one of the biggest timesavers is the ability to search multiple resources from within the Library Catalog. While the standard books came up as expected, so did search results from the *MLA Bibliography* and *PsychInfo* (the two I&A he chose to include). More powerfully, he was able to download MP3 files from the Oral History Project that were included in the results. With these files he could take advantage of the oral tradition and use stories by indigenous people as a primary resource.	▪ Support search across the entire bibliographic information space: catalog, journal articles, digital collections, etc. ▪ Implement metasearch across a wider range and greater number of resources, beyond A&I databases and library catalog.

After a few hours of productive research Santos set up his account so that his search terms were saved for easy retrieval when he searched in other databases. Much of his search terminology and information could be saved in his VLE from which he could build a bibliography and share his resources with his professor.	▪ Deliver bibliographic services where the users are; integrate services such as catalog searching, reserves lists, new book lists, etc into VLE/CMS (virtual learning environment/ course management system), portals, etc. ▪ Work with OCLC to expose all UC holdings to search engines. ▪ Ensure all UC managed content is exposed for harvesting.
In addition to Santos having a productive few hours doing research on his paper, he also gained valuable experience that would directly affect his participation as a student representative on the Library Online Resources Improvement Committee.	▪ As part of our commitment to continuous improvement, task a standing group (existing or newly created) to plan for continuous improvement, environmental scans, user assessments. ▪ Act on plans.

Radical Scenario

User Experience	Library Changes to Implement the Scenario
Keiko, a UC Riverside sophomore, is working on an assignment for her biology class. Her syllabus directed her to the course VLE (Virtual Learning Environment) where she would find the scientific and popular readings she needs to research the scientific arguments used in the debate concerning intelligent design and evolution. Beginning with historical research, Keiko is interested in reading Darwin's original works.	▪ Deliver bibliographic services where the users are; integrate services such as catalog searching, reserves lists, new book lists, etc into VLE/CMS (virtual learning environment/ course management system), portals, etc. ▪ Deliver all bib services via standard protocols such as JSR-168 portlets, Web services, etc. ▪ Expose all UC holdings to search engines directly, push search engines to do better merging, encourage users to restrict to our domain.
The course VLE featured a research "corner" where she was directed to begin her search for monographs in the new OCLC supported catalog.	▪ View UC cataloging as a single enterprise, one that separates metadata store from presentation. ▪ Use OCLC as both the catalog metadata store and the OPAC. ▪ Import bibliographic records to local systems for acquisitions and inventory control only. ▪ Selecting the same record for all UC campuses is challenging when so many record choices exist, catalogers would need to check UC holdings before selecting a record. ▪ Work with OCLC to support UC holdings records. ▪ Build links from OCLC to local circulation and acquisitions functions, including user-initiated features such as recall and RLF paging. ▪ UC-only practice and standards harder to enforce since other catalogers are amending OCLC records.
Keiko immediately finds the original publication, On the Origin of Species (1859), held at UCSF. It is very easy to find texts on evolution; however, intelligent design is a newer concept and gets lost in the teleology mix. For this reason standard metadata doesn't work. Typing the keywords intelligent design into the catalog retrieves mostly engineering and technology titles.	▪ Change how we do metadata – do less of some, more of others; automate more; enrich metadata (ex: with cover art, tables of contents, full text, etc.). ▪ Outsource the majority of MARC record creation. Devote local expertise to specialized metadata creation. ▪ As large portions of our collection become digital, eliminate descriptive metadata and controlled vocabularies for topical subjects for textual items that are self-describing

	■ through search and display of full text. ■ Put extra effort into controlled vocabularies for name, uniform title, date, and place. ■ Add geographic fields with lat/long to all applicable records. ■ Support multiple metadata schemes (ex: MARC/DC, VRA, etc).
Because OCLC has implemented improved discovery tools and incorporated referral systems, TOC, abstracts, notes, bibliographies, etc., the database has a richer pool from which to pull information.	■ Improve discovery tools to get a better result set. ■ Provide intelligent assistance. ■ Crosswalk different metadata schemes. ■ Preference meaningful sections such as TOC, abstracts, notes, bibliographies, etc. for retrieval and relevance ranking.
The results, while still a bit sloppy because of the use of the popular terms "intelligent design" in technology writing, are better at grouping similar concepts and the built in recommender system helps to refer Keiko to titles used by other scholars.	■ Add or change metadata to improve FRBR groupings of related records. ■ Add or change important metadata to improve faceted browsing. ■ Use smart algorithms to allow for faceting on the fly, allowing users to influence the facets being displayed, etc.
What excited Keiko the most was the wide array of current journal articles, pre-prints, blogs, and other resources associated with her topic.	■ Support search across the entire bibliographic information space: catalog, journal articles, digital collections, etc. ■ Create a federated search system by pre-harvesting metadata for the whole space, including catalogs, A&I databases, digital repositories, etc. Consider using Google's harvesting tools, restricting search results to UC-held resources.
Once she chose a number of articles and books she was taken directly to the full-text PDF file without any intervening pop-ups. Additionally, she was informed that the Darwin text she first chose was in micro at UCSF but an easier to read; print version was stored at UCLA's History and Special Collections Cage.	■ Provide direct access to item. ■ UC eLinks takes users to a logical default choice, with option to go back to a list. ■ Allow users to specify turnaround time and tailor options presented based on that. ■ Technology can use RFID to direct users to where item is actually at this moment, rather than just to putative shelving location.
Keiko filled out the pop-up user questionnaire expressing her desire to have the microfilm text transferred to PDF so she could access it immediately, from her dorm computer.	■ Institute our commitment to continuous improvement. ■ Allocate dedicated staff to continuously monitor the environment, conduct user assessments, and highlight new options for UC consideration. ■ Act on plans.

Conclusion

What's information really about? It seems to me there's something direly wrong with the ``Information Economy.'' It's not about data, it's about attention. In a few years you may be able to carry the Library of Congress around in your hip pocket. So? You're never gonna read the Library of Congress. You'll die long before you access one tenth of one percent of it.

What's important --- increasingly important --- is the process by which you figure out what to look at. This is the beginning of the real and true economics of information. Not who owns the books, who prints the books, who has the holdings. The crux here is access, not holdings. And not even access itself, but the signposts that tell you what to access --- what to pay attention to. In the Information Economy everything is plentiful --- except attention.

Bruce Sterling's 1992 speech to the Library Information Technology Association

The Bibliographic Services Task Force was handed the immense opportunity and responsibility of surveying current literature and practices as well as speaking with some experts in the field (Appendix G and Appendix D). We were charged with imagining how to apply the future to how we do business today. We strongly believe that the adoption of these recommendations can lead to dramatic improvements in the user experience and our daily workflow. We look forward to campus conversations during the next year that will build momentum for implementing the recommendations contained in this report which we believe will take the UC Libraries to new levels of excellence.

4

Appendices

Appendix A: BSTF Charge

April 19, 2005

To: John Riemer, (Chair, UCLA), Luc Declerck (UCSD), Amy Kautzman (UCB), Patti Martin (CDL), Terry Ryan (UCLA)

From: Bernie Hurley, SOPAG Chair

Re: **Charge for SOPAG's Bibliographic Services Task Force**

Bibliographic information provides the foundation for all library services provided by the CDL and campus libraries. Over time, a multitude of software applications have evolved to handle the different library services. Because these software applications have been developed to address specific needs, they serve their original purposes well, but do not interoperate as needed in the new shared digital library environment in which we now find ourselves. Various groups have identified problem areas for existing services such as:

- Melvyl and local catalogs: rather than campuses cataloging locally and sending records to be merged in a union catalog, would there be efficiencies and better service to users by using Melvyl as a cataloging utility?

- ERMS: how and for what purpose does the information propagate to other systems such as SFX, including A-Z lists, local catalogs and finding lists? Are there opportunities for efficiency and elimination of duplicate efforts? Would improved discovery services obviate the need for A-Z lists and finding lists?

- Enriched catalog: can the bibliographic data that is in our MARC records be enriched with other data (e.g. ONIX, tables of contents, cover art, etc.) to provide a better user experience? Both Amazon and Google are rich sources of ideas for providing compelling end-user experiences.

- Enhanced resource discovery: what would it take for the discovery tools that rely on UC bibliographic data to have functionality that is common in Amazon.com and Google, such as spell-checking of search terms, and ranking by popularity of the items?

- Design of future systems: how can we build a system that manages the processing of shared print items that complements existing data and systems rather than duplicates it in yet another silo? How can we do the same for a system that manages the digital preservation of UC journals?

- How can we move beyond the limitations of MARC in managing the lifecycle of digital resources (e.g., record information such as book reviews from publishers or notes like "ToC not scanned; irreparably damaged.")

- A single catalog: What would it take for campus libraries to use the union catalog for public access instead of their local catalog?

The University Librarians have directed SOPAG to form a Task Force to rethink how we provide bibliographic services. Therefore, this Bibliographic Services Task Force is being charged to:

1) Inventory the end-user services supported by our bibliographic processing data (e.g., aggregation, discovery, delivery, local and collaborative collection development, collection management, etc.). Identify the middleware, workflow and processes involved in exchanging data between silos of bibliographic information supporting these services. Once the inventory of services and processes is complete, clearly articulate the problem(s) that need to be solved.

2) Develop a vision and design principles for a new bibliographic service environment that states how the underlying bibliographic practices, workflows and technologies can work together more efficiently and flexibly to provide better services to end-users and library staff in a collaborative and shared collections environment (both electronic and print). The vision should provide a compelling story for motivating library staff to do things differently in order to improve user satisfaction. The design principles should address the user experience as well as identify potential architectural models.

3) For services identified in (1), analyze the opportunities to pursue solutions in line with the vision and design principles in (2) and the costs and benefits associated with them.

4) Deliver a report for SOPAG that summarizes your findings in (3) with recommendation on which opportunities should be pursued as high priorities.

5) Develop an implementation road map for those services that SOPAG identifies as offering the most promise to fit the ideals in (2).

The Task Force should send the report identified in (4) to SOPAG by October 3, 2005. Your report will be sent to the ULs, the ULs advisory structure, the CDL and campus libraries for comment. After the comments are collected and reviewed by SOPAG, your Task Force will meet again to develop the implementation roadmaps identified by SOPAG as priorities.

SOPAG thanks you for agreeing to serve on this important task force.

Appendix B: BSTF Membership

John Riemer, (Chair, UCLA)
 Head, Cataloging and Metadata Center

Luc Declerck (UCSD)
 Associate University Librarian, Technology and Technical Services

Amy Kautzman (UCB)
 Head of Research and Collections: Doe/Moffitt Libraries

Patti Martin (CDL)
 Bibliographic Services Manager

Terry Ryan (UCLA)
 Associate University Librarian for the UCLA Electronic Library

Appendix C: BSTF Design Principles

This appendix lists some of the principles that the Task Force believes should guide the re-design of the University of California bibliographic services.

Work Smarter/Rationalize Workflow and Data Flow

The work of both library staff and library systems must be as efficient as possible. We can't waste our scarce resources in unnecessary duplication, complicated algorithms for merging or synchronizing records, or labor-intensive functions that don't add significant value. Our goal should be to reduce the TCO (Total Cost of Ownership) of our collections through streamlining the process of creating and maintaining records.

- Data should be created and maintained in one place. Avoid updating the same record in multiple systems and trying to merge the results.
- Keep data replication to a minimum. Don't load records from one system to another unless absolutely necessary.
- Capitalize on metadata created elsewhere. If metadata already exists, use it as is. Streamline and automate the import of metadata to reduce human intervention to the minimum.
- Not all items (print or electronic resources) receive the same amount of metadata now, nor should they in the future. Select the level of metadata fullness carefully. Only create metadata when its value equals or exceeds its cost.
- Use technology (tools) to reduce the effort to create and maintain metadata
- Focus on being good enough instead of being perfect
- Design and implement systems collaboratively within UC and beyond, whenever possible.

Resuscitate Metadata

Metadata isn't dead but it will be if it doesn't evolve. We need to add a wider variety of metadata and we need to make it work harder. Our systems should do the best possible job in retrieving and presenting results based on whatever metadata we have, the full continuum from minimal metadata all the way to the full text. Scholarly research requires the ability to search in-depth and navigate large results sets, and that success depends on the intelligent use of metadata. Undifferentiated keyword indexing of our enormous information space can result in chaos and noise without the categorization and summarization that can be enabled through quality metadata. Controlled vocabularies and authority control add value when the search and presentation systems take maximum advantage of them.

- Accommodate multiple metadata schemes; don't force conformance to a single metadata schema.

- Enrich bibliographic data with helpful, related information, linked to the broader information universe (e.g., TOC, citations, etc.).
- Support user supplied metadata – for example, user reviews, commentaries, ratings
- Mine existing data and metadata to generate additional metadata (e.g., machine-generated contents notes taken from TOC data)
- Select and adhere to appropriate data standards and best practices, since that will make data usable elsewhere.
- When there's a choice between competing standards; choose the lightest weight that meets the need, avoid complexity that doesn't add value.
- Evaluate controlled vocabulary and authority control approaches and consider where the biggest value lies. For example, perhaps de-emphasize topical subject analysis in favor of more attention to geographic and date analysis.

Provide User-Centered Search Services

We need to provide a suite of bibliographic services with rich functionality and a friendly user experience. Our challenge is to adopt and keep up with standard Internet services that our users increasingly expect while continuing to serve a range of uses, from the quick search for a few relevant materials all the way to in-depth scholarly research.

- Offer options for expanding or improving a search, especially when there are no matching results (e.g., if nothing in local catalog, offer option to repeat search in Mevlyl, WorldCat, Google, etc
- Integrate the discovery of digital and analog resources. Electronic materials are part of the cultural record and must be integrated.
- Provide intelligent assistance to users (e.g., faceted browsing of results, record clustering through FRBR, recommender systems, spelling correction, etc.)
- Support a wide range of users (e.g., make sure that systems are ADA compliant, support all scripts through Unicode compliance as well as search, index, retrieve, display, edit, and sorting functions in multiple languages.
- Reduce the clickstream whenever possible, facilitate self-service, etc.

Get Users to the Content

Our bibliographic services exist to give access to our collections, so efficient, easy-to-use fulfillment services are crucial. Discovery alone is not enough, we must provide the full cycle of Discover-Locate-Request-Deliver. An interaction is not successful until the user has access to the resource itself.

- Never send a user to a dead end, with no options for getting to the content.
- Minimize the number of broken links we present to users.
- For electronic resources, provide immediate links to the content. If the user has identified a specific citation, avoid intermediate screens, the need for subsequent searches, or repeated authentication.
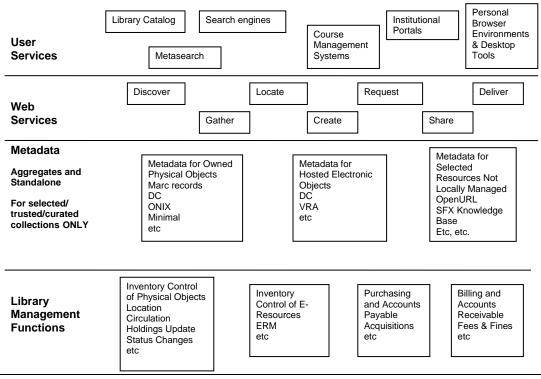
- For print resources, provide immediate information about where to get an item if local, effective delivery options and rapid fulfillment if not local, and Web delivery options for items that can be scanned.

Rethink System Architecture to Focus on Services, not Systems

Separate systems and repositories are still valuable, since they can be optimized to handle specific kinds of data. The services built on top of those systems however, should not be tied to those systems.

Our current user services, such as those provided by our OPACs, are more often than not bolted to the application. Accessing the data stored within our Integrated Library Systems can often only be done through the user interface provided by the vendor. This is very limiting in that it does not allow for the easy remixing of the data or development of alternate user interfaces. Yet, this no longer needs to be the case. With new technologies, such as Web Services, library content can be freed from its application silos. Library content can live separately from the presentation layer and can be offered to others for remixing and repurposing in other environments. This is the approach used by Amazon and Google. They recognized that if the content powering their systems was independent of the presentation and made easily available, additional interfaces would be built by others, thereby extending the search engines' reach and value.

We too need to move away from the traditional systems architecture viewpoint to one where the application becomes defined by the services provided and services accessed by users, as follows:

Service-Oriented Architecture

User Services

Library Catalog | Search engines | Course Management Systems | Institutional Portals | Personal Browser Environments & Desktop Tools

Metasearch

Web Services

Discover | Locate | Request | Deliver

Gather | Create | Share

Metadata

Aggregates and Standalone

For selected/ trusted/curated collections ONLY

Metadata for Owned Physical Objects
Marc records
DC
ONIX
Minimal
etc

Metadata for Hosted Electronic Objects
DC
VRA
etc

Metadata for Selected Resources Not Locally Managed
OpenURL
SFX Knowledge Base
Etc, etc.

Library Management Functions

Inventory Control of Physical Objects
Location
Circulation
Holdings Update
Status Changes
etc

Inventory Control of E-Resources
ERM
etc

Purchasing and Accounts Payable
Acquisitions
etc

Billing and Accounts Receivable
Fees & Fines
etc

We should:

- Separate front-end presentation from back-end processing, so that we can enhance **User Services** and integrate library content into the broader information infrastructure.
- Build user services such as Discover, Gather, Locate, Create, Request, Share, and Deliver as **Web Services** and not system functions tied to particular application silos to enable us to integrate services such as Discover and Request into users' work environment (course web sites, campus portals, etc.).
- In creating federated or metasearching tools across repositories, prefer the model of pre-harvesting **Metadata** where feasible (a la Google, Amazon, Melvyl, etc.) since it allows more control over the user experience and the ability to provide a higher level of service.
- In selecting data interchange standards and protocols to link user services to **Library Management Functions**, give preference to those which have low barriers of adoption, those in use beyond libraries, and those which are likely to be embedded in industry software.
- Plan for an expanding future. We can't know the upper limits for our bibliographic systems, so they should be infinitely scalable.
- Adopt the architectural and behavioral principles of the Web 2.0 era, i.e.,:
    - Provide services, not packaged software
    - Seek to own unique data for competitive advantage
    - Design for user participation, e.g. let users add value by supplementing library-created metadata with user-created metadata
    - Design for remixability. Encourage "mashups" and service recombination
    - Adopt perpetual test model and release new features as soon as they are available
    - Trust users as co-developers, i.e. release new features on a monthly, weekly, or even daily basis, and remove features that are not being adopted by users

Support Continuous Assessment & Improvement

Define outcomes, metrics to measure outcomes, and change/evolve the system to improve outcomes

- Focus on user-centered design, track external research on user behavior and needs.
- Ensure that all metadata is available in a data store that can easily be analyzed, transformed, searched, reported on, etc.
- Ensure that systems are transparent to the designers. Provide documentation, such as entity relationship diagrams, data dictionaries and the like
- Provide a robust reporting capability, crucial to assessment and improvement of our bibliographic systems and of our collections.

Appendix D: BSTF List of Interviewees

Peter Brantley
        Director of Technology, California Digital Library

John Byrum
        Chief of the Regional and Cooperative Cataloging Division, Library of Congress

Priscilla Caplan
        Assistant Director of Digital Library Services, Florida Center for Library Automation

Karen Coyle
        Librarian/Public Intellectual

Tom Delsey
        Consultant, Thomas J. Delsey Consulting

Lorcan Dempsey
        Vice President and Chief Strategist, OCLC

Laine Farley
        Director, Digital Library Services, California Digital Library

Dale Flecker
        Associate Director for Planning and Systems, Harvard University

Brian Kenney
        Editor-in-Chief, School Library Journal

Clifford Lynch
        Director, Coalition for Networked Information (CNI)

Deanna Marcum
        Associate Librarian for Library Services, Library of Congress

Merrilee Profitt
        Program Officer; Research Libraries Group

Andrew Pace
        Head of Systems, North Carolina State University

Roy Tennant
        User Services Architect, Digital Library Services, California Digital Library

Steve Toub
        Web Design Manager, California Digital Library

Appendix E: BSTF Examples to Learn From

The Task Force identified a number of existing systems and prototypes which demonstrate some of the improvements we are recommending. Not one of these systems is an ideal model but all show some element of good practice that UC should emulate.

California State University at San Marcos's RSS Creator
Note how David Walker's prototype allows a library to embed its collection into the user's environment by generating an RSS feed for articles in a licensed journal or newspaper indexed and abstracted in subscription databases searchable via MetaLib, by leveraging information in the SFX Knowledge Base.
http://public.csusm.edu/dwalker/swf/rss-demo.htm

Elsevier's Scirus
An example of federated searching that is generated through metadata harvest and indexing as opposed to metasearching (broadcast searching to each target or source).
http://www.scirus.com/srsapp

Elsevier's Scopus
Another example of enhanced service possible when the system owns or has harvested all of the metadata. Note guided navigation based on dynamic evaluation of metadata in the results set; that is, the ability to refine results by Source Title, Author Name, Year, Document Type, Subject Area. Note also the links to works that cite the article and works within the bibliography.
http://www.info.scopus.com/demo/

Endeca's ProFind
Note especially the guided navigation based on dynamic evaluation of metadata. TLC's OPAC is based on Endeca and North Carolina State University is experimenting with this tool to provide alternative searching within the catalog. Select Endeca Search & Guided Navigation Tour. You must register to view the demo.
http://endeca.com/demos/index.html.

Grokker
Enter a search term into the box and hit the GROK button. Note how the full result set is analyzed and presented in logical subsets, represented graphically.
http://www.grokker.com/

Housing maps
Paul Rade Macher's housing information website which combines Google Maps with Craigslist apartment rental and home purchase data to create an interactive housing search tool.
http://www.housingmaps.com/

National Science Digital Library (NSDL)
Note the attempt to provide specialized portals for different audiences, plus the Explore and Headlines frames.
http://nsdl.org/

North Carolina State University Collection Search
Note the clear presentation of choices, presenting multiple systems to search in a coherent way.
http://www.lib.ncsu.edu/searchcollection/

North Carolina State University new catalog
Built on top of Endeca (see above).   Use the "Keyword search for:" box. Note the faceted browse. Note also Sorting by Relevance and by "most popular" which takes circulation activity for the item into account.
http://www.lib.ncsu.edu/catalog/

OCLC's FAST (Faceted Application of Subject Terminology) Test Databases
Note the attempt to adapt the LCSH with a simplified syntax to create FAST, to retain the very rich vocabulary of LCSH while making the schema easier to understand, control, apply, and use.
http://fast.oclc.org/

OCLC's Fiction Finder Project
Note FRBRization. Link to "search Google" in the record display.
http://fictionfinder.oclc.org/

OCLC's User-contributed Content Pilot (WikiD)
Note the Reviews tab, to allow users to add ratings and reviews, and the Details tab to allow users to transcribe tables of contents and factual notes, to OCLC's Wiki.
Description: http://www.oclc.org/worldcat/open/usercontent/

Personal collection cataloging/management sites
Online services that help people catalog and share their personal book and music collections.
Delicious Monster: http://www.delicious-monster.com/
Flickr: http://www.flickr.com/
LibraryThing: http://www.librarything.com/
Reader 2:   http://reader2.com/
Chain Reading: http://www.chainreading.com/
Connect via Books: http://www.connectviabooks.com/

ProQuest Smart Search
An underlying "engine" that supercharges results by suggesting topics, dates, and publications to help focus your search.
http://www.il.proquest.com/division/pqnext/previews/SmartSearch/

RLG's RedLightGreen
Note especially left-panel faceted browse/guided navigation. Also unobtrusive FRBRization of results.

http://www.redlightgreen.com

Social bookmarking sites
Online services that allow people to bookmark and share their favorite websites.
Del.icio.us: http://del.icio.us/
Shadows: http://www.shadows.com/

Talis' Whisper
A prototype OPAC aimed at showing the discovery of UK based library bibliographic and holdings data using Web 2.0 concepts.
http://research.talis.com/2005/whisper/

University of Buffalo
Using a simple MARCtoXML converter and the TextML indexer from IXIASOFT (http://www.ixiasoft.com), Mark J. Ludwig, Library Systems Manager, indexed the entire collection of two million records in a day and now offers relevance ranking and sort options that are unmatched by most ILS products
http://netcatalog.buffalo.edu (Search page)
http://ublin.lib.buffalo.edu/ub/netcat/start.asp (Background)

University California, Berkeley's Scholar's Box
The Scholar's Box software helps to address important interoperability issues at the intersection of four information technology domains: (1) digital libraries and repositories; (2) educational technologies and learning management systems; (3) web syndication and portal technologies; and (4) desktop applications and structured content authoring tools.
http://iu.berkeley.edu/IU/SB

University of Huddersfield catalogue
Note the recommender features on many records (" People who borrowed this item, also borrowed: ") and floor maps with arrows indicating the likely location of the item.
http://webcat.hud.ac.uk/

University of Southern California's Gandhara
A prototype web-services approach to provide searching across diverse databases including USC's OPAC.
http://gandhara.usc.edu/

University of Pennsylvania PennTags
Note the ability for users to organize and share their bookmarks, and also to tag records from the library catalog (Franklin) and video catalog (VCaT).
http://tags.library.upenn.edu/

WAG the Dog Web Localizer
Note how Ross Singer's service attempts to push content contextually out to users in the places they would think to look by adding a bookmarklet to the browser.
http://rsinger.library.gatech.edu/localizer/localizer.html

Washington State Library's homepage

Note use of geospatial information

Go to: http://secstate.wa.gov/library/

Click on radio button for "Washington Newspapers"

Notice the thumbnail map that instantly appears w/o hitting "Enter"

Click on the map to enlarge it

Click on a county to obtain alphabetical list of titles published within that county.

Wikipedia

Wikipedia is a **free-content encyclopedia**, written collaboratively by people from around the world. The site is a wiki, which means that *anyone* can edit articles simply by clicking on the *edit this page* link.

http://en.wikipedia.org/wiki/Main_Page

Appendix F: BSTF Additional Ideas Considered

Through extensive reading, interviews, and staff comments, the Task Force identified many possible actions for improving bibliographic services and recommended in the report (pg. 8) those which would have the most impact. The other actions we considered are listed here. There are several possible reasons why a particular action was not included as a recommendation in the report:

- o The action offers only incremental improvement, not substantive change
- o The action is premature, though may be useful in the future after new technologies arise and/or there is more evidence of viability
- o The action is not a major recommendation on its own but rather a possible step to consider in implementing one of the broader recommendations in the report
- o The effort to implement the action would outweigh its likely benefit

1. **Add APIs/Web Services for searching and user functions to a single catalog system, enabling each campus to create a local OPAC view**
   Too resource-intensive to implement for the likely payoff. Better to have a single system wide catalog view. Local customized views should aim for a broader information space than just the catalog.

2. **Allow users to add metadata to the permanent record, such as annotations, commentary, reviews, ratings, rankings.**
   It is unclear whether users will be willing to supply useful metadata to the permanent record in enough quantity and quality to be useful. For example, one of the most popular tags is the word "cameraphone", used to tag pictures that people have taken with the products. It may be apt, but it is useless to most of us. We should monitor current experiments in social bookmarking, folksonomies, and the like as applied to bibliographic data, and consider adding these features if they prove valuable.

3. **Abandon classification for full-text electronic materials**
   A possible action item in releasing staff time to enrich metadata in other ways. Since classification is our current hook for doing collection analysis by subject, the loss of functionality may not be worth the gain in staff time.

4. **Assign classification to all materials, including storage facility titles, electronic materials etc.**
   A possible addition to the recommendation to manually enrich metadata since it would allow all materials to be virtually browsed as though they were in the stacks and would support collection analysis tools that currently depend on classification. Not clear, though, that the payoff would be worth the staff time required.

5. **Extend the use of controlled vocabularies beyond the catalog, for example, apply name authority to eScholarship, campus repositories, etc.**

Would extend the benefits of controlled vocabularies to cross-database searching. Not clear that the benefits would justify the cost.

6. **In allocating cataloging resources, invest fewer resources in describing easily discoverable materials (common titles on Amazon, eg) and more resources in describing unique specialized materials (special collections, abstruse languages, etc)**
A possible action item in releasing staff time to enrich metadata in other ways. Not clear that it would apply to sufficient number of materials to yield significant staff savings.

7. **Abandon subject access to serial titles, focus on subject access to serial articles.**
A possible action item in releasing staff time for more impactful metadata work, since scholars rarely search for serial titles by subject. The intent would be to investigate ways to improve subject access to serial articles through automated means rather than manually created metadata. Not clear that the benefit would be great enough to warrant pursuing.

8. **Partner with IS faculty to do research on the economic "value point" for metadata, especially topical subject headings.**
A possible action item to implement the recommendation to consider abandoning controlled vocabulary for topical subjects. Research might help us determine where metadata makes a difference, when the value of metadata equals the cost, and what are the crucial metadata for retrieving material in a scholarly context.

9. **Use separate bibliographic records for each serial format (print, electronic, microform) for efficiencies in processing, and merge at the "work" level for user presentation.**
A possible action step when implementing the recommendations to re-architect the cataloging workflow. Would permit more automatic processing of records. Detracts from user ability to see at a glance all the format choices for a given volume.

10. **Use SFX knowledge base as the description and discovery system for electronic journals, take the records out of the catalog, merge at the "work" level for user presentation.**
A possible action step when implementing the recommendations to re-architect the OPAC. Would allow us to capitalize more on vendor-supplied metadata, redo less in house. Not clear that the data in SFX would be rich enough to allow adequate merging of electronic and print versions at user presentation.

11. **Scale back drastically on creating/enhancing metadata for books in the sciences, standard US/UK publications that come packaged with adequate metadata, etc.**
A possible action step when implementing the recommendations to select the appropriate metadata. Not clear if discipline or place of publication are the most fruitful distinctions to drive the choice of metadata practice.

12. **Invest resources in creating/enhancing metadata ONLY for unique and rare materials for scholarship that distinguish us as a research university library.**
A possible action step when implementing the recommendations to select the appropriate

metadata. Not clear if uniqueness is the most fruitful distinction to drive the choice of metadata practice.

13. Make more use of collection-level records
A possible action item in releasing staff time to enrich metadata in other ways. Not clear that it would apply to sufficient number of materials to yield significant staff savings.

14. Work with publishers to enhance ONIX records with library data prior to distribution
Not clear that we have sufficient leverage with publishers for them to invest in library-directed metadata creation beyond what they already do via CIP.

15. Import MARC records from national libraries for foreign publications.
A possible action item in implementing the recommendation to import metadata whenever it is available. If the new cataloging code, Resource Description and Access (RDA/AACR3), is sufficiently international, records from many more places will be usable within Anglo-American libraries.

16. Import book synopses from publishers, Books in Print, or other sources
A possible action item in implementing the recommendation to add enriched content. Unless it can be acquired easily via Onix feeds, it's not clear that the effort involved would match the benefit.

17. Add or link to evaluative information such as book reviews, and noting when journals are peer reviewed, noting when authors are award winners, etc.
A possible action item in implementing the recommendation to add enriched content, and could help in providing relevance ranking, recommender features, and faceted browsing. Not clear whether we could find enough evaluative information to be impactful. The LC BEAT project experience shows that adding book reviews is not an easy process, may need to license content to get enough to make a difference.

18. Add or link to information about authors, e.g. author home pages, blogs, biographical information, corporate body websites, etc.
A possible action item in implementing the recommendation to add enriched content. Not clear that the effort involved would match the benefit.

19. Incorporate non-Roman vernacular data into metadata
A possible action step in implementing the recommendation to provide better searching for non-Roman materials. Not clear that the benefit would justify expanded efforts in this area. May be more impactful to improve searching, retrieval, and display of non-Roman full-text materials instead of putting additional effort into metadata.

20. Mapping from one metadata scheme to another; map one controlled vocabulary to another and add mapped terms to the record, for alternate forms and richer term sets (e.g. Yahoo categories to LCSH); or create meta-thesaurus (ala ULMS)
Not clear that the benefit would justify the effort. Mapping from natural language or discipline-specific language to controlled vocabulary would be more useful than mapping between controlled vocabularies.

21. Analyze external bibliographies to add "cited by" metadata, to use in recommender systems
A possible action step in implementing the recommendation to add enriched metadata. Not clear that the benefit would equal the significant effort required, not clear that we could add value to what ISI already does.

23. Import links to other systems with relevant and useful information about titles in our collections (e.g., links through Amazon Web Services; search links to indexes that include book reviews; etc)
Adding static links would create a maintenance burden, better to implement dynamic calls as an action step in implementing the recommendation to offer alternative actions for failed or suspect searches.

24. Import (or strategize to federate with) metadata from other cultural heritage organizations, transforming as needed.
Not clear that the benefit would be worth the extensive effort involved, given the current state of metadata in museum and archive collections. We should monitor improvements in this area and consider possible future integration projects.

25. Assign subject terms through linguistic analysis of a broad mix of full text, text of summaries/reviews/TOCs, etc, both controlled vocabulary and keyword terms. Use computational text analysis to discover topical subjects.
Not clear that the benefit would be worth the extensive effort to generate topical subject headings, better to focus on subject facets within large retrieval sets and full-text searching for now.

26. Analyze search logs to identify common search terms to apply as metadata
Unlikely to produce effective additions to metadata. Analysis of logs more likely to be helpful as possible action step to implement the recommendation to offer alternative actions for failed or suspect searches.

27. Extract and normalize date from MARC records, ONIX records, etc.
A possible addition to the recommendation to manually enrich metadata since date searching suffers from the variability and lack of consistency of date encoding, and normalized dates would help with searching, sorting, faceting, etc. Not clear, though, that the payoff would be worth the staff time required.

28. Expand use of record cloning record techniques. Provide clone record capability, so that catalogers can use similar resource descriptions to save time.
Not clear that would represent a substantive improvement in efficiency since most catalog systems already support this feature and the number of records in non-MARC bibliographic systems that would be affected is not known.

29. Provide tools to suggest to equivalent terms in alternate metadata schema. Ex: If DDC call number is present, suggest equivalent LCC number; if LCSH term is present, suggest equivalent MeSH term, etc.

A possible action step in implementing the recommendation to automate metadata creation, but not clear that the tool would enhance cataloger efficiency enough to be worth the effort to build it.

30. **Show the metadata for the *cited* resource, since it might be helpfully analogous for creating metadata for *this* resource.**
Not clear that the benefit would outweigh the effort to implement, since would only be effective if the bibliography of the book were online and actionable, and if the cited items were themselves cataloged.

31. **Suggest subject assignment based on call number and vice versa.**
A possible action step in implementing the recommendation to automate metadata creation, but not clear we could offer additional cataloger efficiency beyond what is provided in LC's *Classification Web* product.

32. **Automatic notification to alert metadata creators when material is deposited or acquired. Notification comes with any metadata that exists at that point, e.g. deposits in the eScholarship Repository.**
A possible action step in implementing the recommendation to manually enrich metadata since it can be very hard to find out when new titles are added to e-resource packages or UC repositories. Could be considered an SDI for technical services, since without this notification, many e-resources would never get metadata assigned. Not clear whether automatic notification could be added for enough packages and repositories to yield sufficient benefit to outweigh the effort required.

33. **Offer option to request expedited delivery for a fee**
A possible action step in implementing the recommendation to provide an "I-want-this" button that is present when the context warrants, with the goal of always offering a fulfillment option. Since many libraries already offer this service, though, it is not clear that this action would yield substantive change.

34. **Minimize number of broken links users encounter that are not explained by a service outage message.**
A good idea, but something that we already try to do. Working to effect direct links to items is more transformative.

35. **Map natural language queries behind the scenes to controlled vocabularies, to support co-location without requiring users to see or understand the sometimes arcane controlled vocabularies**
A possible action step in implementing the recommendation to automate metadata creation, but the state of the art in natural language query analysis may not allow an easy implementation of this idea yet. We also need to decide whether controlled vocabularies for topical subjects are still valuable before exploring this option.

36. **Use discipline-specific ontologies.**
A possible action step in implementing the recommendation to automate metadata creation, since tying our records to discipline-specific ontologies would make our systems more relevant to scholars in that discipline. Using linguistic processing of full text and enriched

vocabulary of works in a discipline could yield a good approximation of the vocabulary of a discipline, that could be used for faceted retrieval. As we gain more access to full text and enriched metadata, we should track the state of the art in linguistic processing, to consider whether this action becomes feasible.

37. **Allow user to set default search settings, such as simple/advanced, preferred type of search, sort order, search limits (language, library, format, etc)**
A possible action step in implementing the recommendation to support customization/personalization, but not clear that this limited customization would provide any significant service improvement.

38. **Allow user to set custom background/color design, as signal of personalized session**
A possible action step in implementing the recommendation to support customization/personalization, but not clear that this limited customization would provide any significant service improvement.

39. **Allow user to save searches to be re-run in the future or automatically**
Most bibliographic systems already have these tools so not clear that improvements in the tools would be transformative. Better to put effort into making the existing services more transparent, and embed them into user environments such as portals and course management systems.

40. **Allow user to design own persistent portal (dashboard)**
Users need to design their own portals at the campus or discipline level, not just within the library. Would be more impactful for us to allow library collections and services to be embedded in other portals.

41. **Allow user to re-open sessions, ebooks, etc where they were last left.**
A useful capability when interacting with full text but not something we can implement at this point given the distributed nature of our current full text systems.

42. **Have user's personal collection of materials and past search behavior influence what they retrieve.**
A possible action step in implementing the recommendations to provide recommender features and relevance ranking, but much more useful to track past search behavior than to try to assess personal collections.

43. **Provide ability for users to "catalog" their own personal collection and to integrate that with resources provided by libraries. (eg, ArtStor)**
Not clear that this is a library service important to provide since other systems and services exist that offer this capability.

44. **Provide ability for end users to share and expose their own personal collection. (eg, LibraryThing)**
Though personal lists and recommendations are frequently shared on Amazon, it is unclear whether scholarly users will be willing to create shared personal bibliographies and collections in enough quantity and quality to be useful. We should monitor current experiments in this area and consider adding this feature if the results warrant.

45. **Suggest other relevant databases for searching based on the search topic or records retrieved.**
A possible action step in implementing the recommendation to provide recommender systems.

46. **Offer contact information for the subject liaison in the library based on the search topic or records retrieved (incorporate the mortal with the portal).**
A possible action step in implementing the recommendation to provide recommender systems.

47. **Offer form fill for user's recently used search terms in the session.**
A possible action step in implementing the recommendation to support personalization/customization, but not clear that users would re-use search terms often enough to make form fill more impactful than a saved search history with the ability to re-execute, which most systems already provide.

48. **Incorporate single sign on and attribute lookup, for automatic authentication and authorization and form fill, for example being able to use a form several times in one session without re-entering name/ID/email address, etc.**
A possible action step in implementing the recommendation to support personalization/customization, but not clear that the technology infrastructure yet exists in UC to achieve it. Must also balance privacy issues.

49. **Identify demands for searching certain subsets of resources and make them available as subsets for user searching. Exs.: ExLibris "logical base" or Innovative "scoping" for serials or for e-resources.**
Though the ability to limit searches to particular formats is clearly valuable, it is less clear that pre-selected subsets would be of use to a sufficient number of users to justify the effort. Better to support on-demand filtering and faceting of large results sets by format.

50. **De-duping before presentation, on content not source**
A possible action step in implementing the recommendation to support searching across the entire bibliographic information space, but straight de-duping could eliminate nuanced differences of importance to scholars. Implementing FRBR concepts, for hierarchical drill-down of related works, is a better choice for scholarly systems.

51. **Common facets across data (shared names, subject terms, places, dates).**
A possible action step in implementing the recommendation to support searching across the entire bibliographic information space, but difficult to implement with the current resources and technological tools. If a pre-harvested metadata store is created, we should assess how viable this action would be in that environment.

52. **Present logical subsets graphically, through topic maps (e.g., see Grokker)**
A possible action step in implementing the recommendation to offer better navigation of large search results, but not clear whether graphical presentation is sufficiently useful to justify the major effort required to provide it. We should monitor current experiments in graphical presentation of bibliographic data, and consider adding this feature if the results

warrant.

53. **Enable library content and services to be easily integrated within a users' personal browser environment, i.e., build browser extensions (such as Peter Binkley's Google Scholar OpenURL extension for Firefox), bookmarklets (such as John Udell 'LibraryLookup' bookmarklet which allows user to lookup library catalogs directly from a browser search box.**
   A possible action step in implementing the recommendation to deliver bibliographic services where the users are, but supporting plug-ins raises the question of how many browsers, and versions of browsers, we will support. Better to put our efforts into providing library services and collections through Web Services to other user environments such as portals and course management systems.

54. **Provide tools that allow users to easily download, export, analyze, manipulate, annotate, and share bibliographic information.**
   Most bibliographic systems already have these tools so not clear that improvements in the tools would be transformative. A better option might be to facilitate Web services linkages to other systems.

55. **Log & analyze user behavior, within privacy safeguards**
   A possible action step to implement the recommendation to institutionalize an ongoing process of identifying and prioritizing improvements to our bibliographic services, though it is unclear how useful doing more with log analysis will be. We already do this, and have probably mined what information we can from the logs given our limited time and resources. A better option might be to track changes in commercial search engines, since they are able to do much more research on logs than we can do.

56. **Log the zero-hit searches for further analysis**
   A possible action step to implement the recommendation to institutionalize an ongoing process of identifying and prioritizing improvements to our bibliographic services, though it is unclear how much benefit comes from this analysis. Many systems have done such analysis and we are unlikely to learn anything we don't already know, though could be used as a benchmark before a change and after, to assess impact.

57. **Provide means of telling us explicitly when users find results unsatisfactory (some supporting systems information is supplied with the communication)**
   A possible action step to implement the recommendation to institutionalize an ongoing process of identifying and prioritizing improvements to our bibliographic services, through such mechanism as a pop-up whenever a search yielded zero results that asks the user to describe what they were doing so we can improve the system. A better solution, though, would be to offer concrete suggestions for further action at that point and possibly track what follow-up action is taken.

58. **Build in a link for user feedback: "Make a comment"**
   A possible action step to implement the recommendation to institutionalize an ongoing process of identifying and prioritizing improvements to our bibliographic services, though not clear how valuable volunteered comments are in driving innovation. Should be paired with more formal assessment.

59. **Conduct periodic focus group interviews re: needs assessment, problem identification, etc.**
A possible action step to implement the recommendation to institutionalize an ongoing process of identifying and prioritizing improvements to our bibliographic services. We do focus groups now, but important to do so more professionally, according to a more defined schedule, and with more commitment to act on the findings.

60. **Conduct usability studies.**
A possible action step to implement the recommendation to institutionalize an ongoing process of identifying and prioritizing improvements to our bibliographic services. As with focus groups, we should improve the professionalism and consistency with which we conduct usability studies.

61. **Build tools to translate search interfaces & help screens for selected languages.** A possible action step to implement the recommendation to provide better searching for non-Roman materials. Not clear, though, that multilingual interfaces are as important as the ability for systems to correctly search and retrieve records in other languages. Though a large number of students and faculty are non-English speaking and UC has goals for greater service to the community and international partners, English is a requirement for success within the UC, our help desk staff only speak English, and reference questions can't reliably be answered except in English.