

November 20, 2006

To: HOTS

From: CAMCIG

Subject: Revised Progress Report:  
Using OCLC as a single cataloging tool/data source

To support the HOTS and SOPAG discussion of the possible use of OCLC as a single cataloging tool for all campuses and as the single data source for the new Melvyl, CAMCIG offers the following information. This progress report is structured:

1. Model description
2. Benefits
3. Requirements
4. Questions needing further investigation
5. Recommendations
6. Appendices:
  - A. Comparison of current OCLC options:
    - FirstSearch Group Catalog
    - Open WorldCat
    - Worldcat.org
  - B. UC Data in the OCLC Database
  - C. Description of a model with three possible OCLC options

The following is what we understand to be the model under consideration by the University Librarians and SOPAG.

## **1. Model description**

UC would use OCLC<sup>1</sup> as UC's shared OPAC in lieu of Melvyl (or of a next-generation Melvyl).

Assuming this model, CAMCIG asserts that one shared ILS or the ten campus ILSs<sup>2</sup> are a key part of UC's bibliographic infrastructure, for a variety of reasons described below. Foremost among them is the notion that UC must maintain a database of record<sup>3</sup>, a role

---

<sup>1</sup> We use "OCLC" to refer broadly to the corporate entity, its main product (the WorldCat database) and the multiple interfaces/products driven from that database.

<sup>2</sup> We use the term "ILS" to refer to the database and the functionality that drive backend staff activity, including acquisitions, serials, maintenance, and circulation functions. We are separating out to consider separately campus web OPAC functionality, although it is also driven by the ILS database. We assume cataloging activity/functionality will happen in OCLC. Whether UC has 10 ILSs or one, ILS functionality is a critical need for UC.

<sup>3</sup> A "database of record" is the single, definitive source for correct and current information. When other databases must be synchronized with it, the data contained therein must not be corrupted. It preserves detailed holding (and transaction) information for asset management, fiduciary and audit purposes, and insurance.

now filled by the campus ILSs. Any change to UC's choice of its database of record must be very carefully considered.

The need for and role of campus web OPACs is unclear. CAMCIG did most of its work under the assumption that there was a continuing need for the local web OPAC because that is the model currently followed by OCLC's product offerings. If development work is possible that would eliminate the need for local web OPACs, then that opens up other options.

OCLC currently offers two products for public access to bibliographic data in the WorldCat database (FirstSearch Group Catalog and the new WorldCat.org), and provides a service (Open WorldCat) through which search engines can link to bibliographic records. These options are described in more detail (including screenshots) in Appendix A. UC would presumably explore a collaborative partnership with OCLC to customize one of these models, or work with OCLC to develop a new product driven by UC needs. A description of three possible options for UC is provided in Appendix C.

## **2. Benefits**

- Leveraging our use of OCLC would avoid significant effort for UC in developing a next-generation Melvyl.
- UC must think outside of its own boundaries in the long-range planning that "rethinking bibliographic services" would entail.
- Eliminating Melvyl will lead to simpler data streams. The current data flow (from OCLC to campus ILSs to Melvyl) could be dramatically changed, and the weekly workload for campus staff to FTP records to Melvyl (and for CDL staff to process and load those records) could be eliminated.
- Eliminating Melvyl will allow campus and CDL staff that maintain Melvyl to refocus their efforts. There will not be a 100% reduction in effort, since any new bibliographic system will require some care and feeding, for example, monitoring contract compliance, problem identification, resolution and escalation, technical discussions, etc.
- Adding in all of UC's bibliographic data and holdings symbols will allow for better collection analysis using the current WorldCat analysis tool, which in turn will yield better collection development.
- More bibliographic data in OCLC contributes to the greater good. If UC needs specific data in OCLC (from vendors and other data suppliers), other libraries will benefit from our effort in consolidating data there. The idea of creating metadata only once throughout the entire library, publisher, and vendor worlds, and maintaining it in only one place, is particularly appealing. A very plausible candidate for that single data store would appear to be OCLC.
- It is to UC's benefit to market its collections (and indeed the University itself) to Web users through more broadly-based Web tools rather than constructing another separate silo. Specific information (availability, detailed holdings, click-through access, etc.) can be provided to authorized users through linking to local ILSs.

CAMCIG notes that transitioning from our current systems to OCLC would require considerable work, both on the part of the campuses and CDL. This will not happen overnight, nor will it be inexpensive.

### 3. Requirements

For this model to succeed, UC has the following requirements that would need to be met:

- A full and accurate reflection of UC's collections. Appendix B provides detail about the 33% of UC's 24 million records (over 8 million records!) which are either not in the OCLC database at all, or which are not shown as held by UC. It is possible to improve this situation (through significant work and expense) by uploading records into OCLC and by projects to identify, match, and add UC's holding symbol(s). (Permission to contribute some record sets to OCLC must be requested from commercial vendors, and agreement to share their data with OCLC is not assured.)<sup>4</sup>
- Display and indexing of customized UC bibliographic data. OCLC's current data model is a "master record," and customization done for local ILS (and Melvyl) access is not retained in the OCLC database. Examples of customized data includes URLs specific to a campus, local access points (e.g., donor names, or collective titles for purchased electronic and shared print packages), and copy-specific information (e.g., autographs, bindings, missing pages).<sup>5</sup>
- A convenient summary/list of campus-specific call numbers and detailed locations. Current OCLC products display a list of campuses. A user must click each campus name to find more detailed information: the specific location, call number, holdings and URL information. Losing this detailed list, which is of particular interest to collection developers, may be an issue for UC.
- Less clicking for users. Current OCLC products require users to click on a link to the campus ILS where detailed information and appropriate links are found. This is one more click than our Melvyl users do now.
- Appropriate handling of non-UC data. More than UC's data is present in Melvyl. Decisions would need to be made about including data from the California Academy of Sciences, California Historical Society, Center for Research Libraries, California State Library, Graduate Theological Union, and Lawrence Berkeley Lab in an OCLC-based UC OPAC.
- Ideally, UC would like to see on-order and in-process data reflected in OCLC, but we note that it is also not now in Melvyl. Only campuses which set OCLC

---

<sup>4</sup> Historically, OCLC has been used as a source of bibliographic data; the concept of a publicly-accessible "world catalog" is relatively recent. The records that campuses use from OCLC are "synchronized" between databases by two techniques: the presence of the OCLC record number in the copied record, and the campus holding symbol "attached" to the OCLC record. The numerous additional sources for bibliographic data (non-OCLC) that campuses have historically used are thus not reflected in the OCLC database. Campuses could undertake projects to improve the synchronicity, but given the magnitude of work, CAMCIG would want a more fully-developed course of action and clear direction before beginning. The expense of cleanup and the ongoing task of keeping the local ILSs in sync with the OCLC database represent significant work.

<sup>5</sup> This requirement might be met if non-RLG UC campuses are allowed to contribute "institutional records" with OCLC, who will maintain and index all copies of local records, following the RLG model..

holdings at order time will have that information available. This may improve with the new OCLC Selection Service under development.

- Circulation status needs to be available to users. Melvyl requires clicking on a link to the campus ILS to display this data, and we assume OCLC would need to do the same for the short term.
- UC would need to record its serial and multivolume set holdings in OCLC. This could be done through an OCLC Local Holdings Project. This is a non-trivial amount of work, although it might be highly beneficial to ILL units internationally.
- SRLF/NRLF as separate locations would need to be explored. This might become important for Shared Print or other collections. Special arrangements may need to be made for these special locations.
- Report writing capabilities are critical, allowing UC staff to create, manipulate, and retrieve primary data, such as creating lists of records based on sophisticated criteria, statistical report generation, and web management reports.
- A cataloging system that can support batch changes is needed to leverage shared maintenance, with a variety of tools to maintain, improve and share data including global update, macros, etc. The sophisticated tools that we have in our local ILSs for database management and maintenance will need to be accessible to catalogers at the OCLC level with OCLC as UC's primary cataloging tool.

#### **4. Questions needing further investigation**

- Are local web OPACs needed? Are they confusing to users and redundant discovery tools?
- How much has UC customized its bibliographic data? (To what extent do we have data discrepancies that are significant from the web discovery point of view?)
- How cost-effective is this model compared with the current Melvyl operation?
- Is there other needed functionality in Melvyl that UC would lose in moving to this model? (e.g., ILL functionality, data of use to selectors)
- Are we certain that Verde and SFX will work with this model? To what extent will UC-based (union) tools such as these be able to function well in both a broader, and a narrower, environment?

#### **5. Recommendations**

1. CAMCIG recommends that UC continue to use campus ILSs as our database(s) of record as we move forward with improvements to our bibliographic services. A single, shared ILS as our database of record is a worthy goal that we endorse that would allow for reduced redundancy and more efficient and shared operations. CAMCIG does not believe that the OCLC database as it presently exists can be UC's database of record. OCLC would need to expand its functionality and its data model—effectively, to become an integrated library system itself. While possible, CAMCIG does not see this happening in the foreseeable future.

2. CAMCIG recommends that UC should continue to utilize other data sources in addition to OCLC. At the present time, OCLC is not complete enough to be UC's single

data source. Appendix B describes a number of large collections<sup>6</sup> to which UC would potentially lose bibliographic access because the records are not in OCLC. While UC could negotiate with their data suppliers to encourage them to submit their records to OCLC, we have no guarantee that they would comply. This would be an interesting leadership role for UC to undertake, but if UC was unable to persuade a supplier to share their metadata with OCLC we could risk eliminating bibliographic access for our users if we were committed to OCLC as a single data source.

3. CAMCIG recommends that further work be undertaken if a promising path with OCLC is identified. More detailed planning and cost analysis will be needed to determine what it would take to get our 8 million “missing” records reflected in OCLC. We should pursue strategies to contribute record packages to OCLC from other data suppliers. UC should re-think whether there is need for a UC-wide holdings symbol in OCLC (especially for Tier 1 electronic resources), and determine the impacts on the Shared Cataloging Program. We will need to be clear about the resources we will need to expend to reflect serial and multivolume set holdings in OCLC. We all need to continue to work toward reduced redundancy and to identify shared efficiencies.

4. CAMCIG recommends that SOPAG engage other UC-wide groups--like Resource Sharing, Collection Development, and HOPS--in exploring whether the option of an OCLC-based OPAC will meet their needs. Groups should also be queried about the need for and role of local web OPACs.

5. CAMCIG recommends that UC continues to explore other options, specifically the single ILS option and the central data file option. Other local data/transaction management systems (i.e., self-standing ILS modules) should be investigated as we become aware of them.

---

<sup>6</sup> Early American Imprints (64,000), CIS records (87,000), ISSR records for datasets, EEBO (90,000), MyiLibrary, ebrary, xRefer, ECCO (130,000), ICPSR (6,000), LION (14,000), ABC-Clio, etc.

## Appendix A. Comparison of current OCLC options

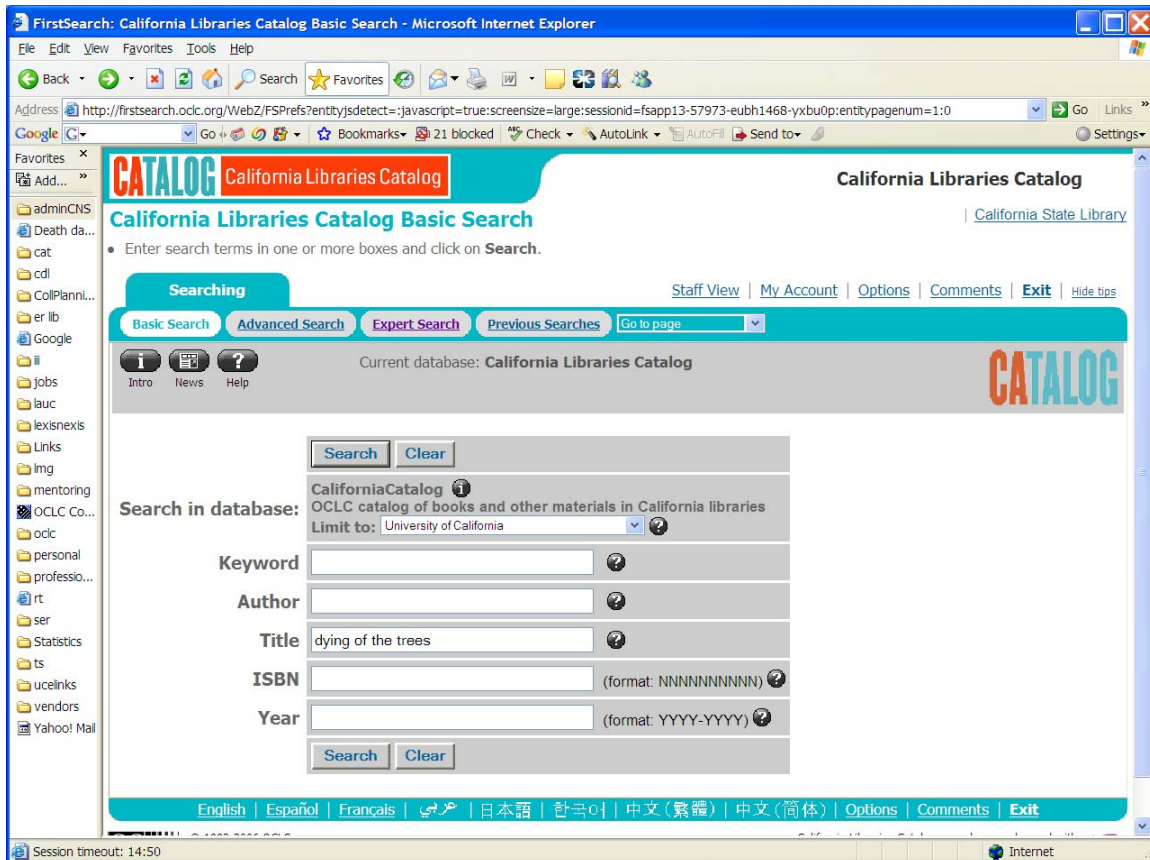
OCLC WorldCat has three “public views”:

- a. FirstSearch Group Catalog
- b. Open WorldCat
- c. Worldcat.org

FirstSearch provides a model via the Group Catalog to be used as UC’s shared OPAC in lieu of Melvyl. They are compared by the diagrams below:

### a. FirstSearch Group Catalog (example: CalCat)

- covers 100% of WorldCat titles
- cross-database searching: simultaneous searching of WorldCat and up to 2 additional databases
- requires authentication unless using guest view
- 50+ searchable keyword or phrase indexes, allowing more focused and predictable searches
- can have sub-view limited to group of libraries by types or subjects
- limit search results: by library/group holdings, language, format, audience, genre, year, etc (most sophisticated)
- detailed display of most MARC fields
- “many interface and functionality customization options available” per <http://www.OCLC.org/worldcat/introduction/comparison.htm>
- updated daily



FirstSearch: California Libraries Catalog List of Records - Microsoft Internet Explorer

Address: <http://firstsearch.oclc.org/WebZ/FSQUERY?format=BI&next=html/records.html:bad=html/records.html:numrecs=10:sessionid=fsapp13-57973-eubh1468-yxbu0p:enttyp:>

Google

California Libraries Catalog

### California Libraries Catalog List of Records

Click on a title to see the detailed record.  
Click on a checkbox to mark a record to be e-mailed or printed in Marked Records.

Searching Results Staff View My Account Options Comments Exit Hide tips

List of Records Detailed Record Marked Records Go to page

Sort Subjects Related Limit E-mail Print Export Help

California Libraries Catalog results for: **ti: dying and ti: trees.**  
Searched in: **University of California (Save Search)**  
Records found: **8** Rank by: **Number of Libraries**  
Search in: **All California Libraries**

Prev 1 Next Search in: University of California Select

Limit results: Any Audience Any Content Any Format Search

- ☐ **The dying of the trees : the pandemic in America's forests /**  
**Author:** Little, Charles E.  
**Publication:** New York, N.Y. : Viking, 1995  
**Document:** English : Book  
**Libraries:** University of California: 7 Worldwide: 1691 **CALIFORNIA STATE LIBR**  
[More Like This: Search for versions with same title and author](#) | [Advanced options...](#)  
[See more details for locating this item](#)
- ☐ **Critical essays on Sylvia Plath /**  
**Author:** Wagner-Martin, Linda.  
**Publication:** Boston, Mass. : G.K. Hall, 1984  
**Document:** English : Book  
**Libraries:** University of California: 8 Worldwide: 1074  
[More Like This: Search for versions with same title and author](#) | [Advanced options...](#)  
[See more details for locating this item](#)

Session timeout: 14:46 Internet

Clicking on the title will yield a list of campuses holding the title, and clicking on the campus will go right into the local web OPAC:

FirstSearch: Libraries that Own Item - Microsoft Internet Explorer

Address: <http://firstsearch.oclc.org/WebZ/FSFETCH?fetchtype=holdingsbrief:next=html/holdings.html:bad=error/badfetch.html:entityholdingsortpage=normal:holdinglimitype=def>

Google

California Libraries Catalog

### Libraries that Own Item

This screen shows libraries that own the item you selected.

Searching Results Staff View My Account Options Comments Exit Hide tips

List of Records Detailed Record Marked Records Go to page

E-mail Print Return Help Current database: California Libraries Catalog Searched in: University of California Total Libraries: 7

Title: The dying of the trees : the pandemic in America's forests Author: Little, Charles E Accession Number: 31661870

[All California Libraries](#) | [Display All Libraries](#)

Libraries with Item: "The dying of the trees : ..."( [Record for Item](#) | [Get This Item](#) )

Location	Library	Code
CA	<b>CALIFORNIA STATE LIBR</b>	CAX
CA	UNIV OF CALIFORNIA, BERKELEY	CUY
CA	UNIV OF CALIFORNIA, DAVIS, SHIELDS LIBR	CUV
CA	UNIV OF CALIFORNIA, LOS ANGELES	CLU
CA	UNIV OF CALIFORNIA, RIVERSIDE	CRU
CA	UNIV OF CALIFORNIA, SAN DIEGO	CUS
CA	UNIV OF CALIFORNIA, SANTA BARBARA	CUT
CA	UNIV OF CALIFORNIA, SANTA CRUZ	CUZ

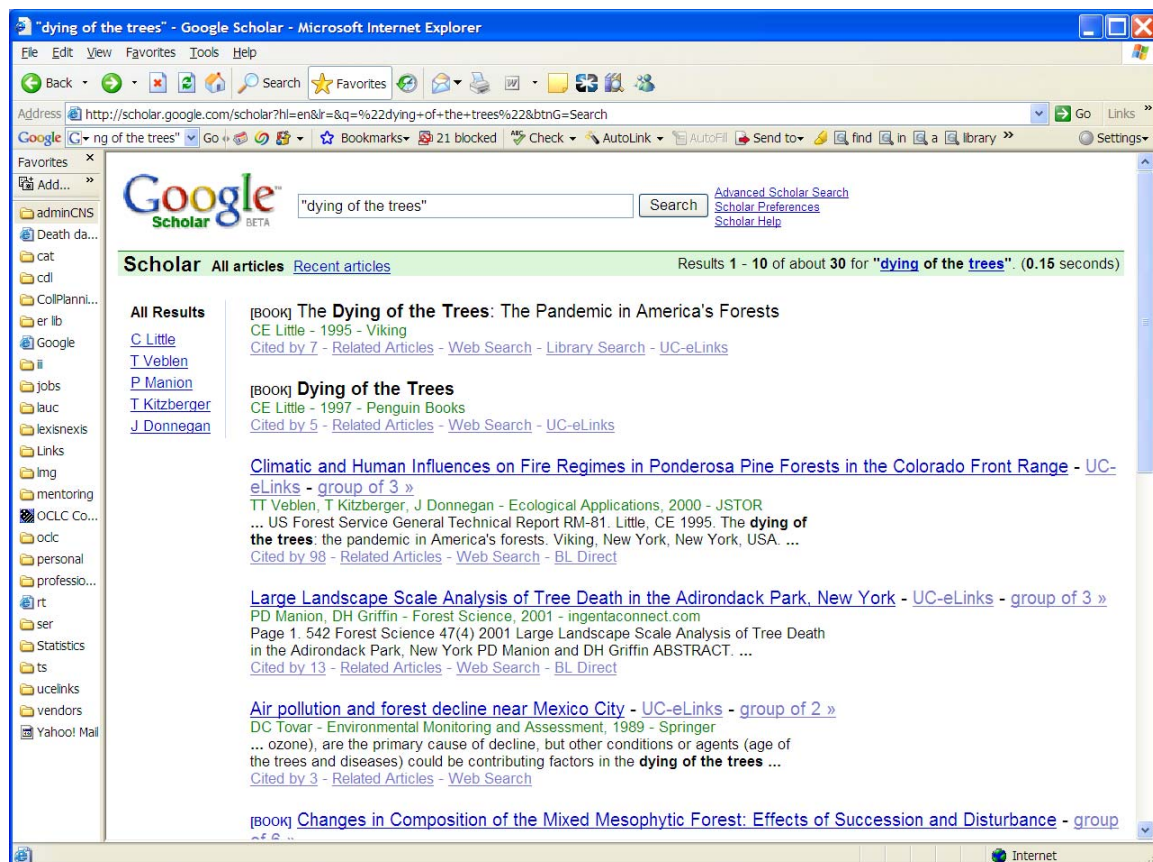
Record for Item: "The dying of the trees : ..."( [Libraries with Item](#) )

Session timeout: 14:56 Internet

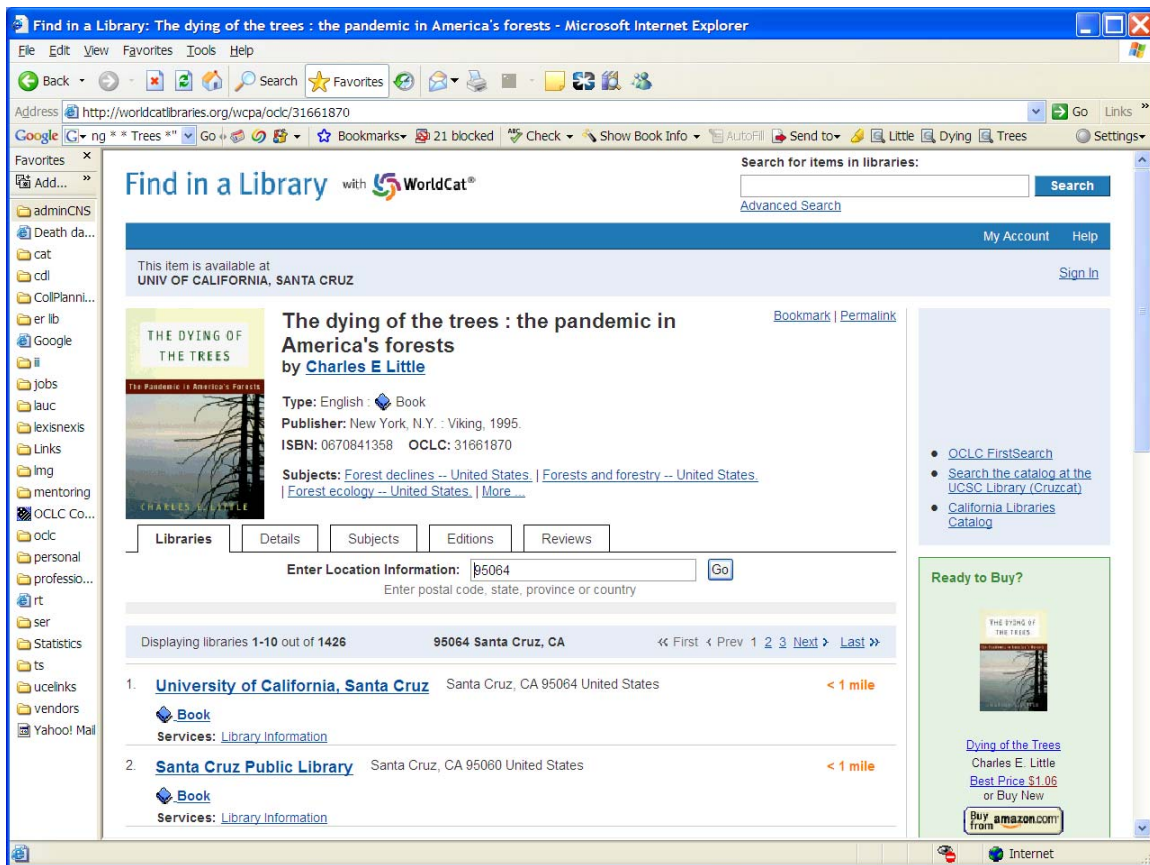


b. Open WorldCat (example: Google Scholar)

- Covers only 5% of WorldCat titles with the most library holdings (will coverage change when number of holding symbols changes?), but indexes also journal articles
- Cross-database searching: Access through links from partner sites
- Great web exposure
- Searching only on partner sites with very limited keyword searchable indexes (author, title, publisher, date, broad subjects)
- Can limit up to three libraries
- No limit search results: search results generated at partner site; library holdings results are limited geographically & narrowed by format
- Brief citation display (title, author, publisher, date, language, format, isbn/issn, subjects)
- Updates varies by web partner
- “Customization limited to displayed links to their web sites, catalogs, virtual reference services and Open URL servers” per OCLC
- Search results include both books and articles and may lead to either library opac or full-text on available platforms for authenticated users. “Cited by”, “Related Articles” and “Web Search” etc. are available.

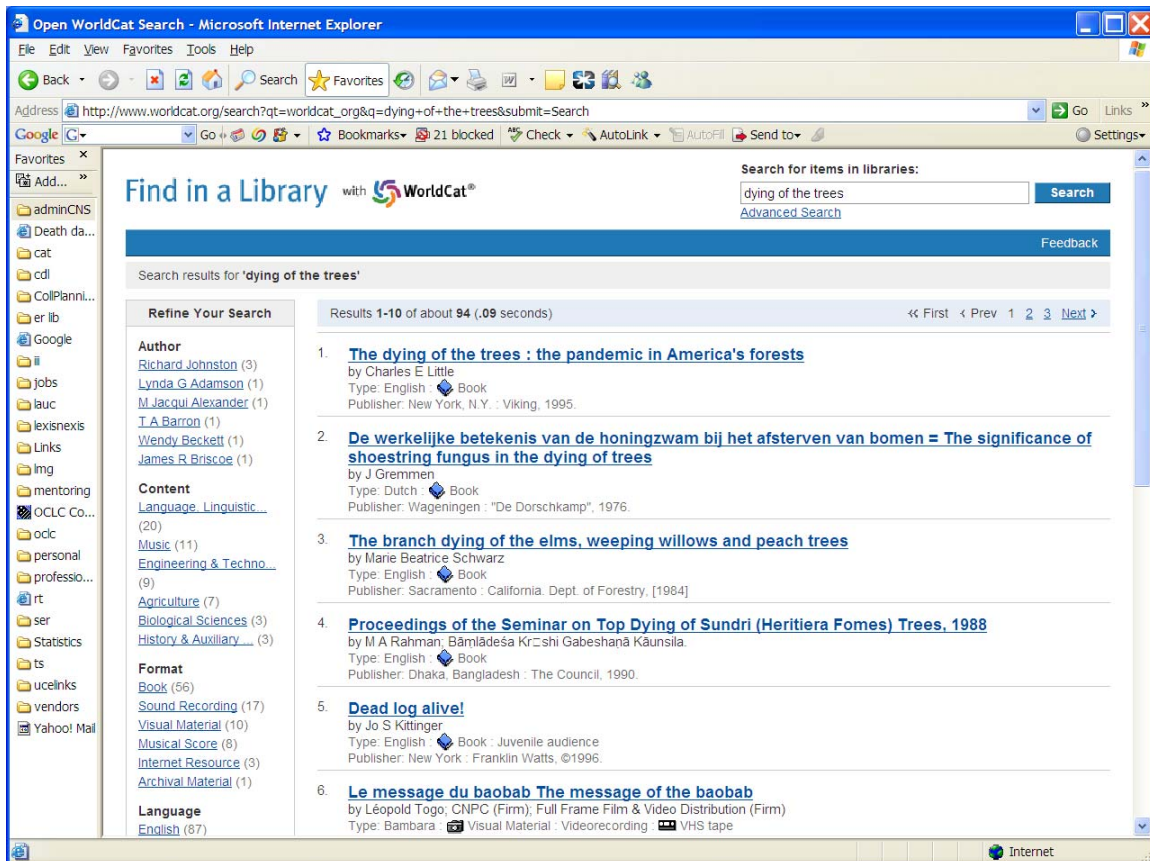
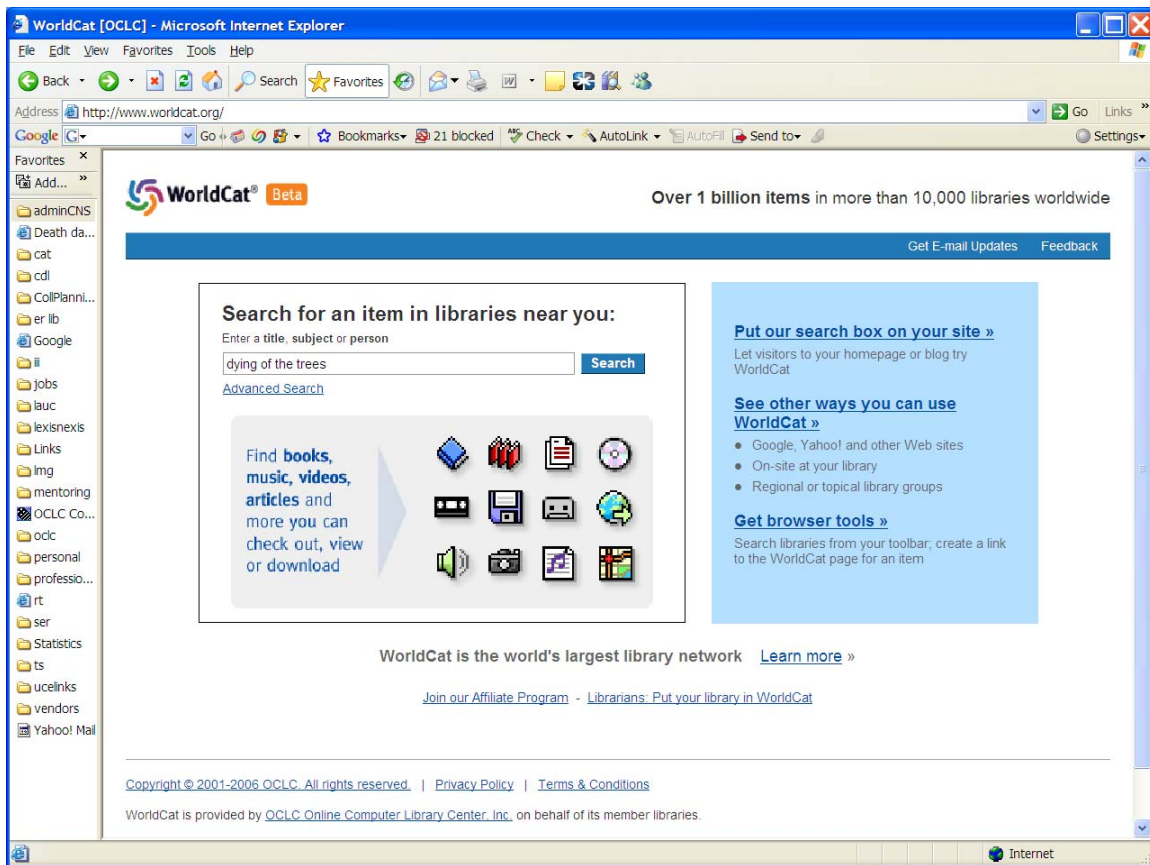






c. Worldcat.org (still an early beta product):

- Covers 100% of WorldCat titles
- Cross-database searching: Multiple database search is planned
- Great web exposure
- Very limited keyword searchable indexes (author, title, publisher, date, broad subjects)
- Cannot limit to group of libraries
- FRBR algorithm grouping; library holdings results are limited geographically & narrowed by format
- Brief citation display (title, author, publisher, date, language, format, isbn/issn, subjects)
- Updates daily
- “Customization limited to displayed links to their web sites, catalogs, virtual reference services and Open URL servers” per OCLC
- Result page above is identical to that for Google Scholar



Clicking on the title will go right into the local web OPAC:

Find in a Library: The dying of the trees : the pandemic in America's forests - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Address [http://www.worldcat.org/odc/31661870&referer=brief\\_results](http://www.worldcat.org/odc/31661870&referer=brief_results)

Google  Search

Find in a Library with WorldCat®

Search for items in libraries:   [Advanced Search](#)

My Account [Help](#)

This item is available at [Sign In](#)

UNIV OF CALIFORNIA, SANTA CRUZ

**The dying of the trees : the pandemic in America's forests** by [Charles E Little](#) [Bookmark](#) [Permalink](#)

Type: English : Book  
Publisher: New York, N.Y. : Viking, 1995.  
ISBN: 0670841358 OCLC: 31661870

Subjects: [Forest declines -- United States](#) | [Forests and forestry -- United States](#) | [Forest ecology -- United States](#) | [More...](#)

Libraries Details Subjects Editions Reviews

Enter Location Information:  95064   
Enter postal code, state, province or country

Displaying libraries 1-10 out of 1426 95064 Santa Cruz, CA << First < Prev 1 2 3 Next > Last >>

1. [University of California, Santa Cruz](#) Santa Cruz, CA 95064 United States < 1 mile  
 [Book](#)  
Services: [Library Information](#)
2. [Santa Cruz Public Library](#) Santa Cruz, CA 95060 United States < 1 mile  
 [Book](#)  
Services: [Library Information](#)

Ready to Buy?

[Dying of the Trees](#)  
Charles E. Little  
[Best Price \\$1.06](#)  
or Buy New  
[Buy from amazon.com](#)

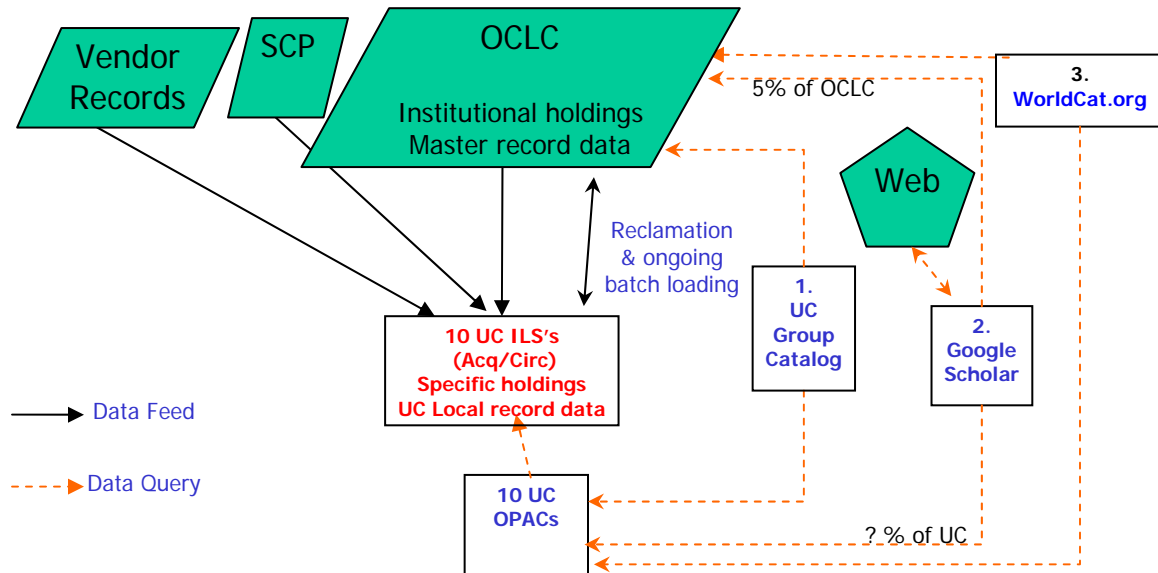
## Appendix B. UC Data in the OCLC Database

<b>Campus</b>	<b>Total Database Size</b>	<b>Records NOT in OCLC</b>	<b>Categories of records not in OCLC</b>
<b>Berkeley</b>	6,170,836 (100%)	1,550,000 (25%)	CJK records MX format records Computer file format records GLADIS record level other than F, R, or B Low-level order order records, circulation-created records, NRLF records, temp cat pool records, some SCP records
<b>Davis</b>	2,365,861 (100%)	1,114,698 (47%)	RLIN records – 608,067 REMARC (Carrollton Press) – 262,254 Early Amer Imprints – 64,543 GPO/Marcive – 44,975 SCP – 58,878 Other – 75,981
<b>Irvine</b>	1,879,407 (100%)	430,871 (23%)	CIS records -- 87,236 Batch loaded records from 1990 -- 89,567 SCP records -- 84,453 Marcive records for e-resources (Documents without Shelves service) -- 62,513 Older marcive records for print California documents records
<b>Los Angeles</b>	5,057,218 (100%)	662,614 (13%)	Monograph (and a few serial records) keyed directly into the database; SCP records for major monograph sets; ISSR (Institute for Social Science Research) records for datasets
			Documents Without Shelves: 68,480

<b>Merced</b>	261,094 (100%)	224,406 (86%)	Vendors (ebrary, netLibrary, MyiLibrary, xRefer, etc.): 44,000 SCP: 72,520 EEBO: 39,406 (92,000 total records will be loaded)
<b>Riverside</b>	1,866,816 (100%)	768,032 (41%)	SCP records Vendor records Hand-keyed records Carrollton Press
<b>San Diego</b>	2,517,728 (100%)	601,242 (24%)	Slide records: 242,911 ECCO: 130,000 EEBO: 96,000 ICPSR: 6,200* Early Amer Imprints: 36,000* LION: 14,000 Carrollton Press: 19,000 Other: 58,00
<b>San Francisco</b>	267,538 (100%)	74,729 (28%)	Print serials SCP electronic serial records
<b>Santa Barbara</b>	2,678,421 (100%)	2,207,693 (82%)	RLIN records (1.4 million; 54%) Marcive (262,000; 10%) GPO (227,000 8%) Early Amer Imprints (36,000; 1%) Congress Hearings (33,000; 1%) Other (177,000; 6%) includes SCP, local originals, etc.
<b>Santa Cruz</b>	1,322,267 (100%)	545,844 (41%)	SCP records (about 17,392 eserials and 72,764 monographs) Marcive Ebook vendors (ebrary, netlibrary, ABC-Clio, XRefer, etc.) (approx. 40,000) In-house creation from publisher md (Lexis-Nexus) (6,000 eserials) Other older records
<b>TOTAL</b>	24,387,186 (100%)	8,180,129 (33%)	

\* UC has contractual agreements that do not allow us to upload these records to OCLC.

## Appendix C. Description of a model with three possible OCLC options:



Like now, each campus would have an ILS with cataloging/circulation/acquisitions/serials modules through which staff would maintain its database of record. In addition to these staff modules, each campus would have a web OPAC (with a locally configured indexing scheme) for users to retrieve information from the local database. Bibliographic data normally comes from OCLC, but use of other data sources is also possible. If data are from non-OCLC sources, they would be batch loaded back into OCLC (if permissible) to ensure consistent access, and to provide synchronization between campus ILSs and OCLC.

### Option 1 (UC Group Catalog or its equivalent):

In lieu of Melvyl, UC would set up one FirstSearch Group Catalog (or its equivalent) web interface for the 10 campuses. This Group Catalog would utilize the WorldCat database of master records with their attached UC campus holding symbols, and the FirstSearch indexing scheme (50+ keyword and phrases indexes). Once a list of titles is retrieved, clicking on something like “University of California” for the selected title will yield the list of campuses holding the title, based on the holding symbols in WorldCat. Clicking on a campus holding symbol will guide users directly to the campus web OPAC for location, call number, circulation and URL data, or OCLC might make local circulation data directly available from campus ILS through possible future developments, using protocols like Z39.50.

In this model, the campus ILSs will be the essential building blocks within the UC/WorldCat bibliographic infrastructure. The Group Catalog might be “good enough” to replace Melvyl, but it cannot replace the campus ILSs. WorldCat cannot provide the needed services for users if there is no campus ILS.

These essential building blocks would change if UC decides to pursue a single ILS. Instead of 10 ILS's, there would be the single ILS. If UC opts for a central database file,



which would be in many ways like Melvyl. The central database file would not change the model too much since it, like Melvyl, does not have a holding symbol in WorldCat.

#### Option 2 (Google Scholar):

This option already exists. Even though it only indexes 5% of OCLC's holdings and uses keyword search, it is a convenient access tool for web users to discover the most popular information. As currently configured, it may serve the ad hoc needs of the "Google Generation", but it cannot replace Melvyl, nor the campus ILSs, which will remain the key part in this UC's bibliographic infrastructure option.

However, it is very tempting to envision turning something like "Google Scholar" into a "Google UC", now that OCLC is going to allow institution records soon. Google Scholar searches both books and articles, employs link resolvers and presents results by geographic distance for books. To incorporate such attractive functionality into a UC union opac, the following developments have to take place:

- Limit search coverage to only UC books and articles in packages/databases etc. that UC is entitled to or has selected.
- Employ protocol to fetch local circulation data for books
- Index UC institution records with local data using OCLC-equivalent 50+ indexes
- Provide searching options: keyword for the "Google generation" and expert search for advanced users
- Provide sub-views and sub-sub-views, e.g.
  - By collections: book collection vs article databases
  - By bibliographic formats: books vs serials
  - By material formats: sound-recordings, video-recordings/films, scores, maps, microforms, electronic, etc.
  - By subjects/genre/categories: medical, law, music, CJK, dissertations, etc.
  - By language
  - By year
  - By campuses: 10 campuses, shared electronic, shared print
  - By campuses & any one of the above appropriate sub-views
  - By campus sub-locations
- Allow selected sub-view to be the default on certain workstations

Course reserves, patron library accounts etc. will have to be dealt with separately

#### Option 3 (worldcat.org):

This option already exists, and will eventually be a very convenient access tool for the general web users. OCLC has intended it to serve the "Google generation" for the web discovery of library collection contents in general. Whether it can replace Melvyl

remains to be seen, but it cannot replace the campus ILSs, which will remain the key part in this UC's bibliographic infrastructure option.