

Proposal for new Ownership and Management of Water Resources Center Archives Library
Submitted by the
College of Natural Resources, University Library (UC Berkeley), and California Digital Library
(UCOP)

Introduction:

Since 1958, the Water Resources Center Archives library (WRCA) has been an important repository for the record of California's water development and use. Programmatic funding from University of California Division of Agriculture and Natural Resources (ANR) through its Water Resources Center (WRC) over the past fifty years has allowed the WRCA to build a unique and valuable collection, and to facilitate access to that collection by maintaining a level of reference service staffing that is unmatched by most specialty collections on the Berkeley campus. This proposal acknowledges the importance and value of these two important yet distinct functions: 1) archiving unique material; and 2) providing support to the research community both on and off the Berkeley campus. While the neither function is dependent upon being physically located on the UC Berkeley campus, the students and faculty at Berkeley have certainly benefitted easy physical access to the collection. In an effort to maintain those benefits for Berkeley's research community, but at the same time recognizing the opportunity to make strategic investments in information technology at this critical point in time that would improve the service provided to the WRCA's wider statewide clientele, this proposal is being jointly submitted by UC Berkeley's College of Natural Resources, University Library, and the California Digital Library (CDL

As ANR closes the WRC as part of a broader effort to restructure and refocus its programs dealing with California water issues and align its staffing more closely with its core areas of expertise, its desire to identify an appropriate administrative home for stewardship of the WRCA collection and continuation of its services is understandable, despite the difficulties of identifying a new academic home in this period of financial stress for the University. Ideally, the unit receiving the collection would be able to offer sufficient space and financial resources to keep the collection intact as a single collection in perpetuity, maintain public access, and provide staffing to maintain and expand

the collection. It is difficult to see how any campus could effectively promise to do so, even with a multi-year promise of financial support that reflected what amounted to a budget cut proportional to what ANR and the campuses have already experienced, especially given the prospect of several more years of, at best, constant nominal funding.

Unfortunately, UC Berkeley is not able to offer financial support that would simply maintain the status quo for the WRCA and its programs, and this proposal has been prepared under the assumption that funding from ANR for the collection during a transition to new management will, necessarily, be at a significantly reduced level. Given the reality of our campus' financial situation and this assumption (which we would be pleased to be informed was misguided), we have tried to identify a path forward that maintains the collection, increases accessibility of water data to the research community, and offers public access to the greatest extent possible.

We feel that it is important to acknowledge, and believe that ANR understands, that any path forward that sustains the collection on the Berkeley campus will require significant transitional funding from ANR. However, we feel that it is realistic to project that at end of a three year transition period, the collection could be maintained and made accessible at an acceptable level independent of ANR funding – and possibly at an improved level with project-oriented funding from ANR's water initiative to support scholarship and services that would be competitive with other demands for support from that source.

Functionally, we propose that the new academic home of WRCA during the transitional period would be the College of Natural Resources, with the possibility of a transition to the UC Berkeley library system in the future. However, because academic research is evolving to become more reliant on data sets and/or documents that are available digitally, we feel that a more robust digital presence is absolutely critical to the future viability of the WRCA regardless of the campus or administrative unit that assumes responsibility for the WRCA collections and programs. In this regard, the CDL adds significantly to the knowledge and resources available on the Berkeley campus – or any campus in the UC system. For the CDL, there are benefits to a significant engagement with the problems associated with the status of the WRCA beyond that of stewardship of this particular collection in that these are symptomatic of those facing many other specialized libraries in the UC system.

Status of Physical Space in O'Brien Hall

The WRCA was founded in 1958 when a special act of California Legislature established the California Water Resources Center to function as a University-wide organized research unit dealing with the state's water resources problems. The WRCA has consistently operated out of space on the Berkeley campus, although the Water Resources Center itself has moved among ANR campuses. As demands for space on campus have grown, the campus has a decreasing ability to offer space to systemwide purposes. There have been several efforts in the past to relocate the archives away from the campus. Currently, an agreement is in place between the College of Engineering and the Water Resources Center to allow the WRCA access to the space until 2013. It is our understanding that the College of Engineering does not wish to renew that MOU and plans are underway for a new use of the space in O'Brien Hall. This situation requires that a proposal from the Berkeley campus must be realistic with respect to the need to identify new space for the collection or integrate the collection into other libraries. There are no options for rent-free space large enough to house the collection on the central campus. Unless rent-free space can be obtained with the assistance of ANR or the College of Engineering can be persuaded to absorb the opportunity cost to its programs of allowing the WRCA to remain in place, the solution at Berkeley will necessarily involve dispersing the collection to appropriate locations. The libraries we have identified as potential recipients for portions of the collection are the Bancroft, Biosciences and Natural Resources Library, the Engineering Library, the Earth Sciences/Map Library, Law Library, Environmental Design Library, and Northern Regional Library Facility (NRLF). Because libraries differ in their accessibility to the public, a key task in the transition period would be developing an appropriate strategy for managing the current collection.

Dispersal of the collection is, realistically, an irreversible action, and will be unpopular with most of the current WRCA clientele – as they have made clear in letters, emails, and meetings regarding the WRCA. Some of their criticisms of dispersing the collection are well-taken. Co-location allows for more efficient access for those who access a diverse array of materials, and has probably increased the effectiveness of the WRCA staff in assisting scholars with finding relevant materials that they might otherwise not have considered. However, unless the funding to provide staffing for a unified collection at historic levels is available from ANR or as the result of significant extramural current-use or endowment fundraising, maintaining the collections in their current location will necessarily

result in significantly reduced access than the clientele have historically experienced. Although it is true that some of the libraries that would be the most appropriate stewards for portions of the collection have restrictions on use to which some of the existing clientele object (e.g., the Bancroft), these restrictions would be balanced by benefits such as maintaining or expanding the hours of access currently enjoyed by WRCA clientele.

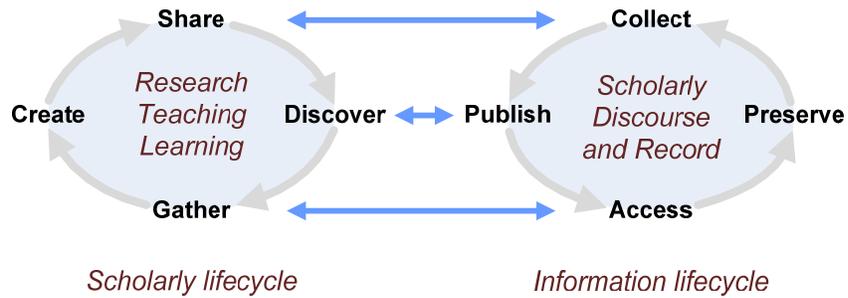
Expanding Digital Access

There is a crisis today in water resources data management due to the volume of hydrologic and watershed data that has been generated and continues to be generated. In addition, questions being raised by policy makers require analysis that incorporates multiple data sets that were typically collected at different spatial and temporal scales and for diverse purposes. CDL has worked with the Berkeley Water Center (BWC) and WRCA over the past year to develop new tools to provide the water research community better access to data sets. This work was originally to be funded by WRC, but when funding for all new projects was withdrawn by WRC, BWC was able to secure funding from Microsoft Research to initiate the work. At that time, the team saw an important role for WRCA as the state's aggregator and harvester of the digital water record, and provider of reference services along with search and discovery tools.

Water-related data sets are varied and have different degrees of reliability. Collected for a variety of purposes, there is an untold wealth of data dispersed around the state. While there are some vehicles for sharing and storing data, none have proven to be comprehensive or entirely effective. Primary barriers to this are variable formats, variable spatial and temporal resolution, and the absolute critical need for metadata or "data about data" that accompanies the information to describe the source of the data, how it was collected, what processing has been done, and previous uses of the data. Those interested in archiving data face two challenges in collecting and managing existing data. First the data are widely dispersed around the state, sometimes in unreliable locations such as individual computers. Secondly, most storage systems are unable to accommodate the flood of data that is generated through remote and real-time sensing devices.

As the WRCA shows, both in its existing form and in the potential outlined in ANR's call for proposals for new management, building and managing a world class collection of scholarly materials requires the integration and long-term care of print, digitized, and born-digital collections,

including published and non-published material of all types and formats. However, to meet the needs of UC students and faculty in the 21st century and beyond, the University must also extend the boundaries of a collection and of core library services to encompass the information lifecycle and to integrate it with the scholarly lifecycle activities of research, teaching and learning.



The California Digital Library (CDL) offers services to manage scholarly materials throughout their lifecycle and to embed and integrate them into research, teaching, and learning. With this integration as a guiding principle, and with its partners and collaborators at the campuses and campus libraries, the CDL builds high quality collections in digital formats; eases discovery and use of materials; publishes and disseminates UC’s scholarship; and curates digital materials for the long-term.

The CDL *builds collections* through **a suite of solution-oriented services** and programs that:

- License 3rd-party content for the UC community (e.g. 34,000 online journals)
- Digitize and host UC content (e.g. 2.5 million books through Google partnership; photos, audio and video from UC and state-wide archives)
- Archive Web resources (500 million web pages)
- Enable researchers to collect and manage their research data

The CDL *eases discovery and use of materials for research, teaching, and learning* through initiatives that:

- Consolidate and integrate discovery tools with the online lives of scholars and students (e.g. Google integration; Web2.0 tools; *Melvyl*; *UC-eLinks* ;*UCVerse*)
- Describe and index the scholarly record in both old and new forms (e.g. the addition of datasets, web-sites, working papers, etc. to *Melvyl* and *Next Generation Melvyl pilot project*; *Online Archive of California*)
- Leverage a single copy or item for systemwide use (centralized *Shared Print Collections*; *Request* and interlibrary loan tools)

The CDL *disseminates UC's scholarship* through programs that:

- Provide a flexible publishing platform for all stages of the scholarly lifecycle, from dataset through preprints, conference proceedings, seminars, journals, and books (e.g. *eScholarship*; *UCPubs*; *DataOne*)
- Thematically group and integrate materials harvested from diverse sources (*Calisphere*; *UCVerse*)
- Enable editorial and reader-commentary overlays on previously published material (*Calisphere*, *UCVerse*, annotation services)

The CDL *enables curation of digital materials for the long term* through initiatives that:

- Apply or combine a set of micro-services to create trusted, enduring repositories (e.g. *Digital Preservation Repository*; *Web Archiving Service*; local application of curation components)
- Create standards, policies, and tools that enable long-term stewardship of materials (e.g. persistent identification; integrity and fidelity checking; format migration; *JHOVE2*, *UDFR*)
- Address the particular challenge of access, use, and endurance of scholarly and scientific data (e.g. *DataCite*; *DataONE*, *DAX*)

WRCA – the Next Generation

CDL services currently offer core support for the management and use of WRCA materials, which can be:

- collected, preserved, and accessed in the California Water Districts Web Archive (<http://webarchives.cdlib.org/a/CAWaterDistricts>);
- digitized through our partnerships with Google and the Internet Archive, and managed via HathiTrust (<http://hathitrust.org>) or eScholarship (<http://www.escholarship.org/uc/wrc>);
- discovered in the Next Generation Melvyl Catalog pilot (melvyl.worldcat.org) and Google Scholar;
- delivered via UC-eLinks ([http://www.cdlib.org/services/d2d/UC-eLinks /](http://www.cdlib.org/services/d2d/UC-eLinks/));
- accessed and read, among other articles and technical reports, in eScholarship (<http://www.escholarship.org/uc/wrc>);
- discovered, and in certain cases viewed, in the Online Archive of California (<http://www.oac.cdlib.org/institutions/UC+Berkeley::Water+Resources+Center+Archives>).

Through collaboration and partnership the organizational host of WRCA can leverage additional CDL initiatives into a discipline-specific suite of services that supports scholarly and community interaction with current and future water-resource related materials throughout their lifecycle. Doing so would extend the “archive” as an active resource for all stages of scholarly investigation, dissemination, and use. Among many possibilities are:

- Digitization of current analog materials for Internet-wide discovery and use
- Curation and publication of gray literature, datasets, and new forms of discourse found in blogs, websites, “scholar-sourced” commentary, etc.
- Long-term care and curation of digitized and digital materials
- Systematic addition of new scholarly genres and content related to water resource issues

Combined with the curation of the WRCA collection per se, these services would allow the WRCA to become a Next Generation scholarly resource, well-positioned to support – by providing and caring for a full range of scholarly outputs - emerging forms of scholarship such as virtual research environments and collaboratories, interdisciplinary research which includes water-resources as a component, and data curated as a “living” product of scholarship. The CDL stands ready to provide more details about these services, demonstrate production services and next generation prototypes, and join specific planning and action initiatives when and as appropriate.

Transition Actions and Budget

If UC Berkeley is selected to accept and maintain the WRCA library collection, a significant effort to plan for the transitions will be initiated. We anticipate that full transition will take three years to complete. Because the space in O’Brien is still available for at least two years, the library can still remain open to the public (with reduced service), while a more detailed transition plan is developed. The funding requested to support this transition is \$200,000/year for three years, or a total of \$600,000. We anticipate that these funds will be augmented significantly with funds raised from other sources. The transition plan would take approximate twelve to 18 months to develop and 18-24 months to carry out.

Primary tasks associated with creating a transition plan are as follows:

1. Maintain access to collection in 413 O’Brien with reduced hours
2. Create a transition task force, including representatives from DANR and the WRCA advisory board

3. Work with WRCA staff to develop a staffing plan
4. Initiate user needs survey of the Archives users asking for input on the relative importance of types of documents
5. Collect information on methods and costs of digitization, including working with Google and Hathi Trust
6. Communicate with Foundations and other funding sources to assess possible additional funds
7. Communicate with all UC libraries to assess their interest assuming portions of the collection and discuss coordination of the collection growth systemwide.
8. Complete a detailed transition plan based on information collected

The \$200,000 annual expenditure is broken down by personnel, acquisitions, collections management (scanning, removal of print journals), and outreach.

Budget

| Year | 1 | 2 | 3 |
|---|-----------|-----------|-----------|
| Personnel | \$100,000 | \$80,000 | \$40,000 |
| Acquisitions | \$15,000 | \$15,000 | \$5,000 |
| Management/Digitization/Relocation of collection | \$80,000 | \$100,000 | \$150,000 |
| Outreach | \$5,000 | \$5,000 | \$5,000 |
| Total | \$200,000 | \$200,000 | \$200,000 |

Appendix 1: Service Glossary

Descriptions of CDL applications, services, and tools mentioned above; for more information about these and additional services please contact the CDL at cdl@www.cdlib.org.

Calisphere <<http://calisphere.cdlib.org>> offers educators, students, and the public direct online access to more than 200,000 digitized primary source materials from educational and cultural institutions across California, organized into thematic collections.

Curation Micro-services < <http://www.cdlib.org/services/uc3/curation/>> comprise an approach to digital curation based on devolving curation function into a set of independent, but interoperable, services that embody curation values and strategies. Since each of the services is small and self-contained, they are collectively easier to develop, deploy, maintain, and enhance. Although the individual services are narrowly scoped, the complex function needed for effective curation emerges from the strategic combination of individual services.

DataCite <<http://datacite.org/>> is an international consortium, of which the CDL is a founding member, providing solutions for enhanced access to scientific research data, increasing acceptance of research data as legitimate, citable contributions to the scholarly record, and supporting data archiving for verification and re-purposing for future study.

DataONE <<https://dataone.org/>> is an NSF-funded DataNet initiative focused on ensuring the preservation and access to multi-scale, multi-disciplinary, and multi-national scientific data about life on Earth and its environment, including biological data from the genome to the ecosystem, and environmental data from atmospheric, ecological, hydrological, and oceanographic sources.

DAX is a project funded by the Moore Foundation developing best practice recommendations for the publication, archiving, and citation of technical appendices providing crucial expanded documentation of how data are analyzed, synthesized, and processed to reach scientific conclusions presented in the scholarly literature.

Digital Preservation Repository (DPR) <<http://www.preserve.cdlib.org>> serves the stewardship mission of the UC libraries by providing a single shared solution for the preservation, management, and controlled dissemination of digital collections that support research, teaching, and learning. The repository provides a set of self-service interfaces that the libraries use to deposit and manage digital objects, relieving individual libraries of the burden of creating and maintaining custom digital repositories.

eScholarship <<http://www.escholarship.org>> is the CDL Publishing Group's open access scholarly publishing platform, providing digital publishing services to the University of California and delivering a dynamic research platform to scholars worldwide

Google Partnership < <http://www.cdlib.org/services/collections/massdig/>> is one of a suite of mass digitization projects through which the UC Libraries are digitizing millions of books from their collections (another is with the Internet Archive). These projects expand the UC Libraries' ability to give faculty, students and the public access to information and support our exploration of new service models

JHOVE2 < <https://confluence.ucop.edu/display/JHOVE2Info/Home>> is a mechanism for the automated examination of digital objects to determine their significant properties and implications of those properties.

Licensed Resource Program < <http://www.cdlib.org/services/collections/licensed/index.html>> provides electronic journals, databases, ebooks and other e-resources licensed by CDL on behalf of and in coordination with the ten UC campuses.

Melvyl <<http://melvyl.cdlib.org>> contains records for materials (books, journals, movies, maps, music scores and recordings, computer files, dissertations, government documents, etc.) held by the libraries of the ten UC campuses, and select partners. The database contains over 33,000,000 records, and most campuses update their holdings on a weekly basis.

Next Generation Melvyl Pilot Project <<http://melvyl.worldcat.org>> is a pilot version of a replacement for the current Melvyl Catalog, featuring a single interface that searches holdings in all UC libraries, books digitized by Google, materials from the Online Archive of California and eScholarship, as well as those of libraries around the world through OCLC's WorldCat. It also searches for article references in many major disciplinary databases.

Online Archive of California <http://oac.cdlib.org> provides free public access to detailed descriptions of primary resource collections — and more than 170,000 images and documents digitized from them — maintained by more than 150 contributing institutions including libraries, special collections, archives, historical societies, and museums throughout California and collections maintained by the 10 University of California (UC) campuses.

Request [see UC-eLinks]

Shared Print Collections The University of California Libraries' Shared Print collections consist of information resources jointly purchased or electively contributed by the libraries. Such resources are collectively governed and managed by the University Librarians for the purpose of maximizing access to the widest audience of current and future members of the UC community.

UC-eLinks connects scholars directly with articles and ebooks by providing an easy way to move from an article or book citation to the actual publication. It also lets users check to see whether an item is available in a UC library or to request that an item be delivered to them by interlibrary loan.

UCPubS <http://www.escholarship.org/publish_ucpubs.html> is a collaboration between the CDL's Publishing Group and UC Press, that offers UC departments, research units, and publishing

programs the flexibility to provide free digital access to their book projects while retaining the ability to sell print copies.

UCverse <<http://ucverse.universityofcalifornia.edu>> is a multimedia portal project that will enable anyone anywhere to virtually “attend” UC, the world’s largest public university, and participate in discussions of topical issues, focused initially on climate change.

Unified Digital Format Registry (UDFR) is a project organized by a loose consortium of national and academic libraries and archives, with the aim on consolidating the existing [PRONOM](#) and [GDFR](#) format registries. CDL staff serve on governance and technical committees.

Web Archiving Service (WAS) <<http://was.cdlib.org>> provides you with the tools to easily capture, analyze and preserve web content. The web archives created with WAS can be searched by keyword or browsed by site. WAS web archives are free to the general public, and WAS capture tools are available by subscription.