

Introduction to the California Digital Library

Daniel Greenstein

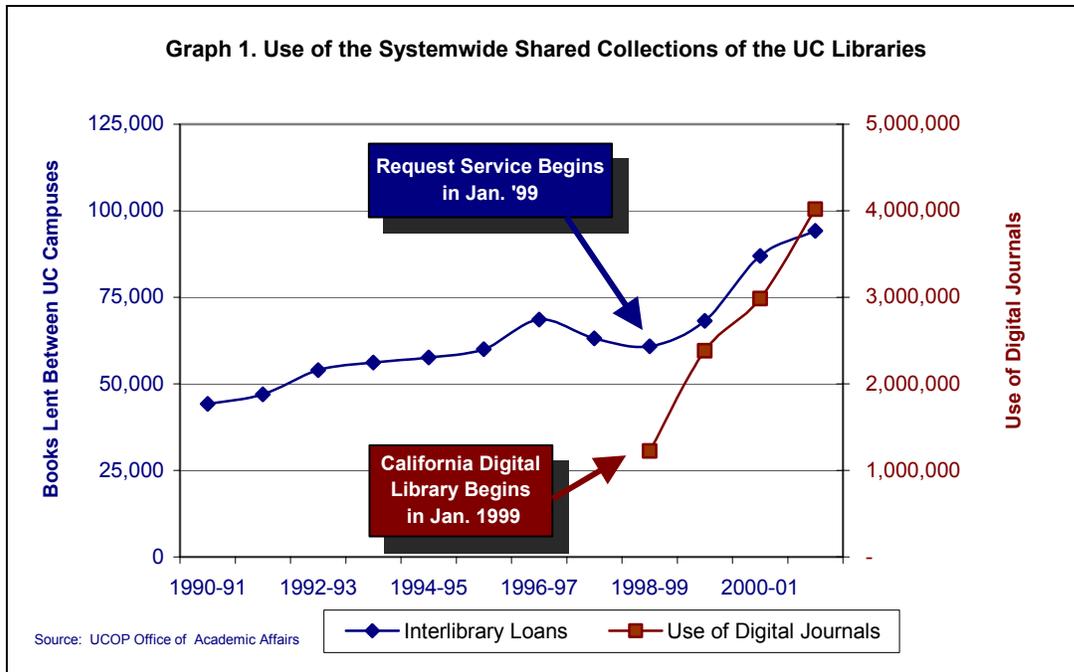
October 23, 2002

The CDL was established in 1997 as part of a planning process that seeks to guide the university libraries through a transition from a campus-based and print-centered service model to one that blends print and digital information and more effectively leverages the shared resources and capabilities of the UC system. Its aim is to develop the infrastructure that supports this transition and enable campus libraries to provide access to high-quality collections and services as appropriate to a research university of UC's distinction.

Major components of the CDL's infrastructure and a brief indication of their benefits are listed below

1. By integrating information about their holdings, UC libraries encourage access to and use of their respective collections as if they were part of a single virtual collection. Melvyl is a union catalog that combines in a single place, information about the monographic, serial, and other largely bibliographic holdings of the University of California and other major regional libraries. Patron-initiated request (or simply, Request) is a service that is built on the basis of Melvyl with which patrons can initiate interlibrary loan of a holding that is known to Melvyl but not available on the patron's campus. As shown in Graph 1 the use of Melvyl alongside an efficient courier service that has been developed by the UC libraries has doubled the number of interlibrary loan requests over a ten-year period, with the greatest growth occurring since the introduction of the Request service. We are expecting further upward progress with interlibrary loan with the introduction of an electronic document delivery service through which requests may be scanned (digitized) and delivered to a patron's desktop rather than sent via the courier service. While Melvyl provides access to bibliographic materials, the Online Archive of California affords online information about the holdings of over 7,000 UC and other archives and special collections. This infrastructure:

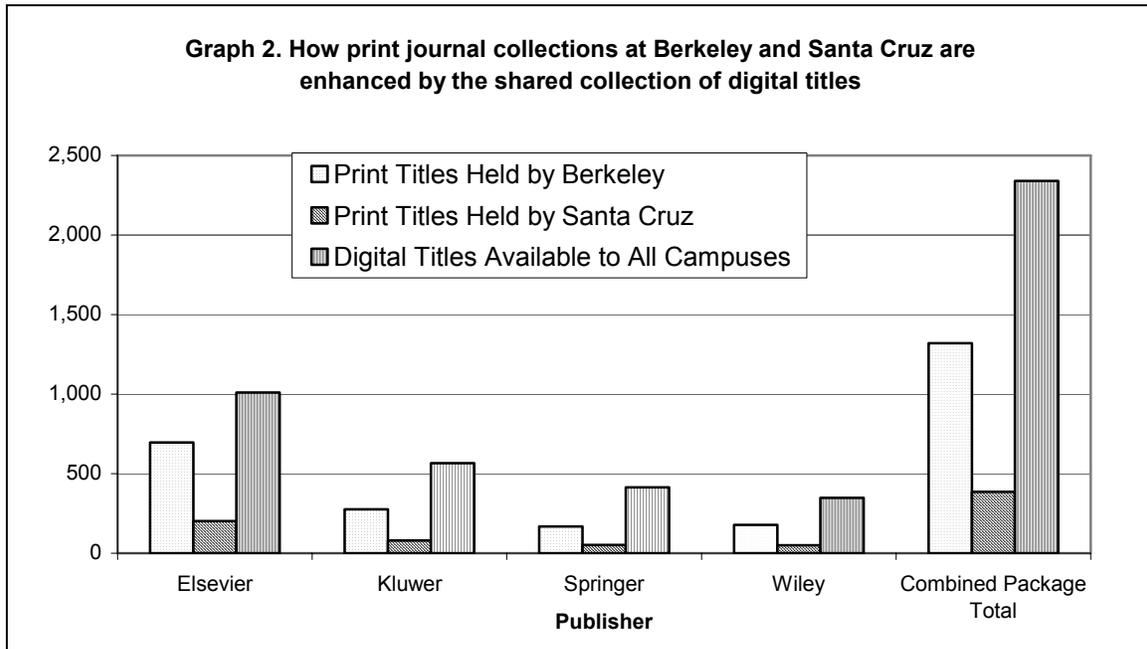
- encourages use of campus collections as if they are single uniform collection and
- encourages collection development staff at campus libraries to make collection development decisions in light of what already exists in the system as a whole (we have some evidence that this actually happens).



2. Development of shared collection. The University of California libraries act as a single entity in developing a shared collection of online digital material. At present the shared digital collection comprises more than 7,000 journal titles, 200 abstracting and indexing (and other) databases, as well as other material. The shared electronic collection does three things.

- It expands and enhances collections available to UC campus libraries. Graph 2 shows how the journal holdings of UC campus libraries large and small benefit are extended considerably by the shared collection.
- The shared collection also drives down the costs associated with acquisition of commercial electronic content. With respect to electronic journals, for example, the University spent approximately \$16,139,888 for access to 2,515 periodical titles from eleven of the top publishers. The cost covers all nine campuses. Had all of the campuses subscribed to all of the titles on their own, the additional cost to UC for a single year would have been \$24,354,018. By building a shared collection, UC effectively achieves a 58% discount from the average subscription cost and a gain of 13,791 additional subscriptions systemwide. Similar savings have been achieved with databases. For example, in a recent ProQuest deal, the University received a discount of approximately \$953,513 on a package whose total cost was \$1,740,245 and which included three years of waived access fees for all campuses.
- The shared collection finally enables UC libraries to think systematically about and provide for persistent access to electronic journal titles and database holdings (e.g. digital preservation), and also about rationalizing print holdings where electronic equivalents exist. At present, the UC libraries are investigating

mechanisms for developing physical collections of last resort for print journals that also exist electronically in the shared collection.

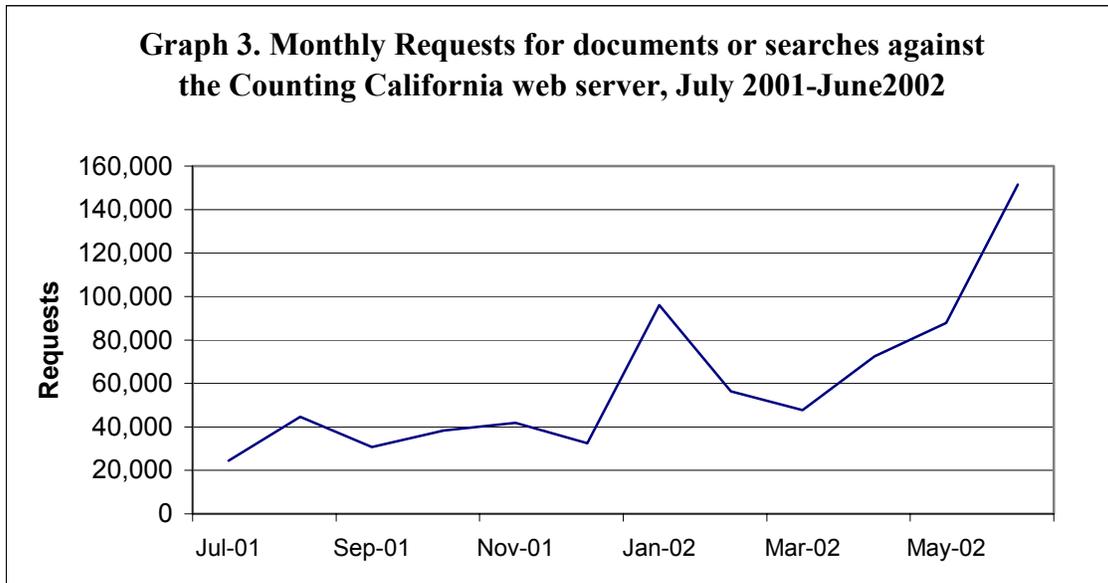


3. Services, tools, and sources of expertise that lower the cost to campus libraries of developing high-quality online information environments and (to those who wish to), persistent, and interoperable digital collections and finding aids.

Tools and services are listed below.

- An SFX server enables libraries to build online services that link citations (e.g. in A&I databases), to the underlying full text content (e.g. in online journals). The CDL built an SFX server in order to enable linking across content in the University's shared collection. The service, known as UC e-links, has proved enormously popular with faculty who want to see it extended to electronic content that is not part of the shared collection (campuses maintain their own collections of electronic information over and above that available to them from the shared collection). Accordingly, UC libraries are leveraging the CDL's development work to extend linking service to their local electronic collections. The CDL's PURL server (enables us to keep track of where digital objects – encoded texts, digital images, etc. – are located and how they are named) has similarly been leveraged. It is used by a number of campuses to keep track of their digital collections.
- Digital repositories allow UC libraries (and others) to persistently manage digital information (including digitally reformatted objects made from special and archived collections, scholarly working papers and e-prints, etc.) and to make it accessible to the widest possible community. At present, CDL provides digital repository services for statistical data from US and California state governments of relevance to the citizens of California (Counting California); for selected digitally reformatted materials from holdings in UC archives and special

collections (Online Archive of California and California Cultures); and for digitally reformatted images of works in architecture and the visual art (Museums in the Online Archive of California). The success of these repositories is evident in their growth (UC libraries are rapidly contributing digital information to them), and in their use. The Online Archive of California, for example, has evolved rapidly to include well over 100,000 digital objects (texts, images, etc). The Counting California service extends to some 3,000 statistical datasets and has attracted considerable use as shown in Graph 3.

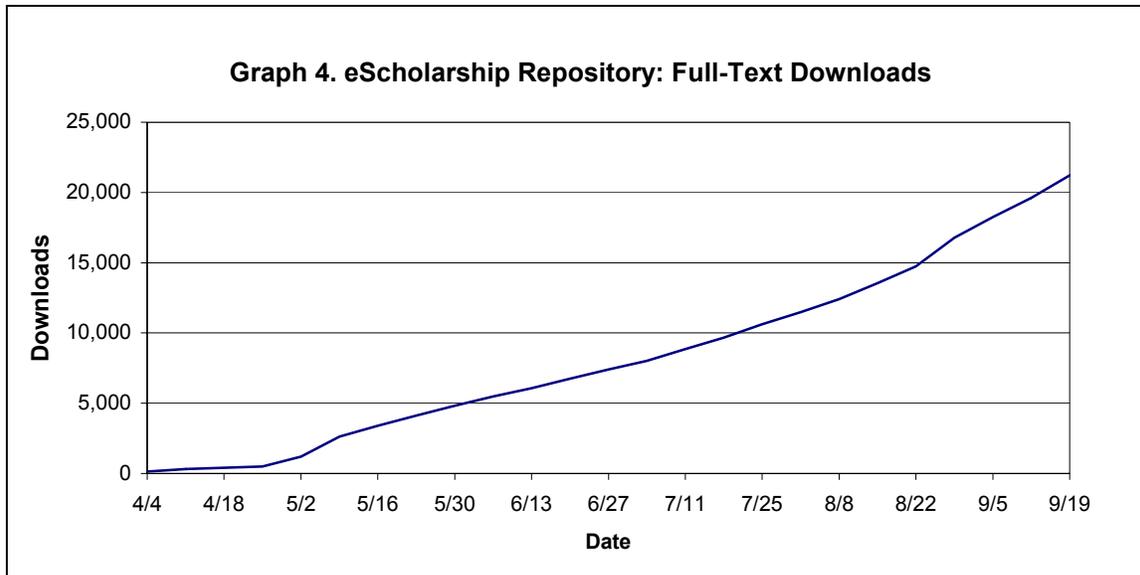


Expertise is a by-product of the CDL’s work– one that is itself of significant value to individual UC libraries. It is available in a variety of forms, for example as:

- ❑ licensing templates that campus libraries can use in negotiating acquisition of or licenses for commercial electronic content that is purchased locally;
- ❑ agreed data creation practices that lower barriers to UC libraries who wish to develop online finding aids for the archive and special collections or to digitally reformat items in those collections; and
- ❑ methods for assessing use and usability of online services helps UC libraries evaluate and iteratively extend and improve their digital library initiatives.

4. Support for new modes of scholarly communication. The CDL’s eScholarship program provides infrastructure and leadership in order to stimulate and facilitate innovation in scholarly communication. The program is very active and has assembled repository through which UC faculty can persistently manage and distribute pre-prints, working papers, articles, monographs, scholarly journals, and other products of their research and teaching. It also enables and facilitates online peer review. One aim of the repository initiative has been to encourage faculty to seek alternatives to commercial publishing vehicles that, because of their pricing structure or their protracted publishing process, have become impediments to scholarly communication. The eScholarship repository has

been well received. The repository has been available since May 2002 when CDL began systematically soliciting interest from organized research units, campuses, and departments. The repository has been enthusiastically received by the organized research units, academic departments and centers that have been approached during the preliminary roll-out phase. To date some 30 organizations from six campuses have signed up to use the repository and contributed nearly 700 articles and working papers to it. Despite its novelty, and its relatively small collection, use of the papers in the repository has also grown at a phenomenal rate as demonstrated in Graph 4, which shows downloads of papers in the eScholarship repository since launch.



5. Applied research. CDL works collaboratively with UC libraries, computer science departments and schools of library and information science to conduct applied research into areas that are critical to the development of the infrastructure described above, and to the development of high-quality campus based online collections and services. Its participation in national and international associations and initiatives also ensures that digital library development activities at UC are informed by and contribute to such efforts as they evolve globally. The applied research agenda is a long one and effectively supplements the typically scarce resources that individual campus libraries can apply to this essential activity. Historically, it has also been innovative and brought national acclaim to numerous UC-based initiatives, some of which are listed below.

Through its collaborative digital library efforts (including those managed by the Department of Library Automation – a unit subsumed by the CDL in 1997), UC has developed (in some cases invented) numerous methods and applications that have gone on to become national, international, and industry standards, including:

- Z39.50, a network search and retrieval protocol that enables distributed online databases to be searched as if they were a uniform database
- SFX, a protocol for linking between citations (e.g. in an abstract and index database) and full text of the cited article (e.g. in a journal)

- ❑ Encoded Archival Descriptions (EAD), the international standard for creating and encoding online finding aids for archive and special collections
- ❑ Metadata Encoding and Transmission Scheme (METS), a mechanism for recording information about a digital object as required for its persistent management and interchange

Leveraging their buying power in the market place for commercial electronic information, the UC libraries have also had a significant influence in the market place for commercial electronic information. It is responsible, for example, for the emergence of the “preservation clause” as a standard feature in licenses between publishers and libraries. The “preservation clause” ensures that libraries have perpetual access to the digital information to which they subscribe.