



# World Library and Information Congress: 69th IFLA General Conference and Council

1-9 August 2003, Berlin

---

**Code Number:** 065-E  
**Meeting:** 141. Library and Information Science Journals  
**Simultaneous Interpretation:** -

## **eScholarship at the University of California: sustainable innovation for Open Access**

**Carol Ann Hughes**

Ph.D., University of California Irvine  
Irvine, USA

---

### ***ABSTRACT***

*This presentation describes the history of the University of California eScholarship program, a joint effort of the University of California Libraries in collaboration with the California Digital Library. It discusses the context that gave rise to the creation of the eScholarship repository, the logistical issues involved in setting up a multi-campus persistent repository for scholarly output, and future issues to be addressed in developing experimental reconfigurations of the components of scholarly communication in collaboration with communities of scholars.*

### **INTRODUCTION**

Over the next decade, a significant challenge for research universities is to influence and facilitate development of a financially sustainable model for managing scholarly information. Scholars are slowly seeking the opportunity to develop strategic innovations in scholarship – production of information as well as use and access - that match their needs with the opportunities created by digital technologies. Such faculty innovations promise the likeliest means to sustain and enhance the international scholarly communication system.

Institutional, and in the case of the University of California (UC) supra-institutional, repositories have recently become a key strategic initiative on many campuses as a way of supporting the production, dissemination, and preservation of new scholarship that fully exploits the promise of digital technologies. In this paper I will use the definition developed by Cliff Lynch, Executive Director of the Coalition for Networked

Information (CNI), that “an institutional repository is a set of services that a university offers to the members of its community for the management and dissemination of digital materials created by the institution and its community members.”(1)

The international scholarly communication system still relies heavily on commercial publishing for the management and dissemination of scholarly content. However, scholarly publishing still focuses largely on models derived from the print world in which information is disseminated at lengthy intervals in “branded bundles” called journals or books. Most publishers have severe limitations on the ancillary data, tables, etc., which can be included in any given publication. Certainly little is known about the intentions of the bundle producers, i.e. scholarly publishers, relative to permanent preservation and access to the content of this electronic scholarship.

Institutional repositories offer an opportunity to support innovation in scholarship that can not be represented fully in print. They have the added features of nearly unlimited accommodation for ancillary data, and a much greater opportunity for presenting the work of a scholar in a timely manner and in the context of its relationship to other work. There is no absolute requirement that a repository allow “open access” but without that foundational principle a repository can not hope to fulfill its promise as a means of improved dissemination of scholarly work. And robust access and preservation of scholarly work is assured at the institutional level by those who have traditionally had responsibility for the intellectual assets of the institution, i.e. librarians, archivists, and information technologists which is a benefit that “self-archiving” enthusiasts can not guarantee.

Some also claim that institutional repositories are a means by which universities can begin to leverage significant change in scholarly publishing by providing a competitive outlet for peer-reviewed scholarship. As Lynch notes in the article cited above, universities have historically had a passive relationship to scholarly publication even though they invest heavily in supporting the creation and dissemination of the published content. Many others have written about the serials pricing crisis and the need to free scholarly publication from proprietary interests (and proprietary servers.) An excellent example is the work by Mary Case in her overview of the Scholarly Publishing and Academic Resources Coalition (SPARC) initiative of the Association of Research Libraries.(2) It is true that repositories may be used as infrastructure for restructuring the current economics of scholarly publishing. However Lynch warns against encumbering an institutional repository with the kind of gate-keeping policies for admitting materials that would be necessary to significantly challenge more traditional outlets. He sees the real value in repositories as that of encouraging innovation and presentation of academic work of all types grounded in the local community of scholars.

The University of California Libraries, in collaboration with the California Digital Library (CDL), has recently launched the eScholarship Repository, an online repository of multi-institutional faculty research and scholarly output, as its response to pressures for innovation in publishing and scholarly communication. The eScholarship program offers a UC-supported infrastructure that both meets the scholarly community's needs for open access to peer-reviewed material and is extensible to other types of projects. It is a

hybrid model in that it has a low threshold for accepting materials into the repository that are both peer-reviewed and non-peer reviewed. But it also has the goal of influencing the scholarly publishing system by providing tools for peer-reviewed books and journals that may even be migrated from commercial publishers into the repository.

It costs nothing for UC units to join. Papers are uploaded by a repository system administrator located in the sponsoring unit rather than by individual faculty members. Faculty retains copyright for their papers and may post or publish them in other locations as well. The Repository provides persistent access to working papers and makes them easily discoverable.

This presentation describes the history of the eScholarship program, the logistical issues involved in setting up a multi-campus persistent repository for scholarly output, and future issues to be addressed in developing experimental components of scholarly communication in collaboration with scholars themselves.

### **THIEVERY vs. OPEN ACCESS**

During the launch of D-SPACE at MIT<sup>1</sup> last November, James Boyle, professor of law at Duke University, mentioned that he had recently read a list of the “top ten thieves in the new information economy” promulgated by the content industry. The Number One category of “thief” in the new information economy was librarians. They joined the ranks of other new types of thieves, such as people who fast forward through commercials, or worse, who go to another room of the house during commercials. These people were thieves (in the estimation of content providers) because they were comfortable with and acted in support of the concept of free content. After discussing trends in intellectual property law that are moving toward a perfect system of pricing and control at the level of the individual user for content of all kinds, whether or not it has commercial value, Boyle discussed the need for a new initiative in society. He stated the need to allow authors to express content in a way that allows it to be shared easily if that is the author’s intent. He went on to discuss the Creative Commons initiative.<sup>2</sup> But he could just have easily gone on to talk about open access initiatives that allow scholars a technical infrastructure to offer their intellectual works to the world of scholarly communication, of which the eScholarship program at the University of California system is a powerful example.

Cliff Lynch was the next speaker at the MIT event. He made the point that research repositories and e-print repositories have advanced quite unevenly across disciplines and none at this time easily accommodate multidisciplinary content. Progress on a scale sufficient to impact the scholarly communication system at large is dependent upon a model that takes better advantage of the faculty culture at an institutional level. Lynch went on to discuss the need for data archiving at the local level, but again he could just have easily gone on to talk about the eScholarship program at the University of California.

What makes the eScholarship program and its Repository such an intriguing model? It is interesting because it combines a mixture of freely available peer-reviewed and non-peer-reviewed content contributed by faculty at ten of the world's premiere research institutions with the attention to robust technological infrastructure and persistent access/preservation provided by the California Digital Library. It is interesting because it encompasses an academic community of shared purpose across disparate campuses and a variety of disciplines. It is interesting because it includes local campuses in the process of building content for the world – it is, in other words, a grassroots effort. Plus, and maybe most importantly, it is simple for faculty to participate.

## **HISTORY OF eSCHOLARSHIP**

According to the UC website,<sup>3</sup> eScholarship was launched as a result of the University of California (UC) Library Planning and Action Initiative, which was charged "to identify changes ... required to ensure the continued scholarly and economic vitality of UC's libraries." The LPAI task force declared in its 1998 report: "The present system of journal publication no longer meets faculty needs to distribute information quickly and effectively. Commercial journals are too slow to publish new scientific information, their peer review processes are perceived as cumbersome, and prices limit distribution to a few relatively wealthy institutions in developed countries. . . . Publishers of digital information are placing restrictions on its distribution and use while they have yet to establish methods to archive this information and ensure that it will be readily available to future scholars and students. To capture and distribute effectively the fruits of the knowledge developed by UC faculty requires new forms of scholarly and scientific communication."

The University of California, one of the largest institutions of research in the world, at that time employed faculty who served as the senior editors of approximately 12.5% of the world's most prestigious scholarly publications. Therefore suggesting that an entirely new system of distribution of scholarly content was "required" was a somewhat bold step into a potentially contentious arena. The California Digital Library (CDL) which initially acted primarily as host to MELVYL, the UC union bibliographic catalog, and as coordinator of licensing arrangements for UC campuses, was now to include in its mission the creation of an infrastructure to support the comprehensive management of scholarly information from production to dissemination and long-term access and preservation. This also was an adventurous step. To ensure that their new agenda could be well served, CDL set up an eScholarship program in July 2000 to move forward on building a digitally-based scholarly communication system.

The overall goals of the eScholarship program became the development of an infrastructure that:

- Facilitates the expressed mutual interests of the University, its faculty, and the broader scholarly community;
- Leverages the formidable capabilities and strengths of the University of California in order to provide effective national leadership in this area; and

- Supports and extends experimental reconfigurations of the components of scholarly communication by communities of scholars themselves.

The eScholarship program has several distinct parts. One highly publicized section is the UC International and Area Studies (UCIAS) Digital Collection of edited volumes and monographs. These are peer-reviewed materials in which the individual chapters are made available as soon as they are ready for distribution. Another effort, *Dermatology Online Journal*, was created to explore the educational potential of distributed hypermedia served over the Web. This early project was intended as a model of what can be done and not a prototype. However, CDL has just launched an infrastructure that is intended to be extensible for the support of new or migrated electronic journals. The criteria at this time for e-journals are that they be free and available to any Internet user; digital only, without a print component; sponsored by a UC research unit, center, or department; and able to use the currently available technical infrastructure, which supports PDFs along with associated content (images, Excel, etc.)

Another effort that is utilizing the distinct advantages of the Web is the Electronic Cultural Atlas Initiative (ECAI) *ePublication Series*. This initiative provides stable, long-term access to peer reviewed, map-based digital scholarship in history and the humanities. ECAI publications include a text component, a web-based map, and a fully interactive downloadable map.

A fourth programmatic initiative of eScholarship is the University of California Press eScholarship Editions. These monographs are made available through a partnership between the University of California Press and eScholarship. More than 750 titles are available now, 400 of which are open to the general public. The rest, because of UC Press restrictions, are available only to UC, faculty, students, and staff. Eventually there will be over 1500 monographs online.

In April 2002, the eScholarship program launched the eScholarship Repository<sup>4</sup> as the fifth major initiative. It offers faculty “a central location for depositing research or scholarly output deemed appropriate by their participating University of California research unit, center, or department, including working papers and pre-publication scholarship. The repository provides persistent access to working papers and makes them easily discoverable.” The repository is the only open access initiative that is actively recruiting new content from the general UC population.

Currently the Repository contains 1,500 papers from more than 100 UC institutes, departments, research units and centers. As of April 2003, users had logged 100,000 full-text downloads since the Repository’s inception. Over 95% of the downloads are coming from outside the UC system.

## **OPEN CONTENT INITIATIVES**

At the joint conference on institutional repositories held in October 2002, sponsored by the Association of Research Libraries, the Coalition for Networked Information, and

SPARC, the Scholarly Publishing and Academic Resources Coalition, Jim Neal, Vice President and University Librarian at Columbia University gave some intriguing opening remarks. He likened the current state of repository infrastructure as the “napsterization” of scholarly communication. Indeed, encouraging peer-to-peer sharing is a strong impetus for the development of repositories. He noted that the urge to publish is a natural one in our academic cultures, but that we had a responsibility to preserve the public stake in research information and that it is a public policy opportunity for universities and their libraries to preserve information in the public domain.

Joe Branin, Director of Libraries at Ohio State University, reported at this conference that OSU has integrated their repository initiative within the entire “knowledge management” system of the Library. However Rochester University’s Susan Gibbons, cautioned against making the library a crucial component in the deposit process for a repository. Catherine Candee, Director of Scholarly Communication Initiatives at CDL, gave a presentation that explained that the eScholarship Repository has some attributes of both systems, as will be explained below.

Ann Wolpert, Director of Libraries at MIT, also made some important observations during that conference. She noted that librarians have to keep in mind that faculty have their own ideas about metadata (we don’t own the only schemes that matter). One crucial word of advice from her was that librarians should approach building repositories as a partnership with faculty, not as an academic service for the faculty. It seems clear that local campus libraries have much more to offer in electronic publishing initiatives than merely ingesting what scholars offer. Yet what should be the parameters for making the partnership effective as a publicly reliable content storehouse? This is one of the early issues with which that the eScholarship program grappled.

## **UC eSCHOLARSHIP REPOSITORY CONTENT**

The Repository accepts content from faculty, staff or students in the UC system. Content can range from working papers to peer-reviewed series. However there must be a sponsoring body that approves all content – CDL staff do not assume responsibility for accepting or denying content submissions into the repository. The sponsoring body decides which content is admitted into the Repository and it chooses a system administrator to do the actual uploading of materials. The system administrator is trained to use software from the Berkeley Electronic Press (bepress)<sup>5</sup> for content management. Sponsoring bodies must assure that all materials, including illustrations and ancillary materials, have appropriate legal clearance for public dissemination.

This model of content solicitation, i.e. there must be a sponsoring body to ensure copyright compliance, etc., has by default privileged certain categories of scholarly output. Working papers and publication series already located within the structure of an academic department or institute have been the first to be contributed. Learning objects, electronic dissertations, and other types of individual scholarship that do not naturally have a “sponsoring unit” have not gravitated into the Repository. Some faculty in the humanities have challenged the “sponsoring unit” approach as being too focused on

supporting only products of collaboration rather than individual scholarship. Currently the only procedure for an individual scholar to participate in the Repository is for the department or some other bureaucratic unit to act as the sponsoring unit on her behalf. There may be only one person contributing to that “series,” but she still needs an official sponsoring unit to assure that all appropriate copyright permissions and technical procedures are followed. It remains to be seen whether the Repository will change its policy to encourage the contribution of learning objects and other faculty materials. To date we have not heard many of our community asking for inclusion of these types of objects, but that can change as more local discussions take place.

The process for soliciting content has changed in recent months. During the first wave, established research units were identified and contacted by staff at the California Digital Libraries and Berkeley Electronic Press (bepress), with whom the CDL contracted to help roll out the repository. Now that the wave of “early adopters” has been accommodated, a new method that provides more local contact and encouragement has been implemented.

Solicitation of new participants is now reliant on support from local campus librarians as well as word of mouth from other faculty. Each UC campus has an eScholarship liaison with whom CDL and bepress staff communicate about the state of discussions with faculty in various stages of repository adoption. Campus liaisons are trained in the basic parameters of the system and are charged to follow up with faculty who may be considering joining the repository. It is also expected that campus libraries may host an event locally or provide other customized outreach efforts to engender interest in eScholarship activities.

As you might imagine, campus librarians prefer to be involved in any service that is so innovative and so important to faculty as a new way of publishing their research content. The initial model of having staff from the UC system or bepress, the commercial contractor, contacting local faculty did not allow local librarians to assume a key role as partners with their faculty in reconceptualizing scholarly communication. The current model allows for much more local contact with faculty who are interested in the Repository. This task has been easily integrated into regular faculty liaison duties and it presents no additional workload. It is just another opportunity to promote the role of the campus library in faculty concerns.

The requirement for a UC sponsoring unit notwithstanding, authorship of content in the Repository is not restricted to UC affiliates. For example, a unit may use the repository to post papers from a conference they sponsored, which includes some UC authors and many from other institutions. All that is required is that the sponsoring UC unit approves all content for their area of the repository.

Data sets and other non-textual materials are welcome in the Repository, but they must be accompanied by some kind of textual commentary if they are not part of an article. If the dataset has any special technical features, the capacity of the bepress software must be able to handle it.

Peer-reviewed series have a few unique policies. Material in the eScholarship Repository that has been peer reviewed carries a label which acts as a link to a description of the peer-review process employed for that unit's submissions. Papers in a peer-reviewed series may not be revised or removed after they are posted. Additionally all peer-reviewed series include an author review step before addition to the repository is completed.

Papers that are not peer-reviewed may be revised or even removed upon request of the author, but a citation will always remain with a note about the withdrawal. Multiple versions of a paper may be posted for any period of time. If agreements with commercial publishers allow, papers may remain in the repository after publication for non-commercial dissemination.

Authors retain the copyright for all papers posted in the repository. Escholarship's agreement is nonexclusive so the author is free to reuse the content elsewhere, either in the same form or in revised form.

UC units can deposit any content into the eScholarship Repository that meets the guidelines and conforms to technical and policy requirements. Material that does not conform to repository guidelines, or that does not come through UC administrative units, needs to be deposited elsewhere. Thus some UC campuses, including UC Irvine, are currently discussing the idea of creating a campus-based repository for learning objects and items that are not sponsored by an administrative unit. Conversations have not gone very far at UC Irvine because of our severely depressed budget situation. But we have a great partnership with UCI's Network and Academic Computing Services unit upon which to build when the financial picture becomes clearer.

## **TECHNICAL ENVIRONMENT**

The eScholarship Repository is currently set up to accept and serve PDF documents. There is a conversion server that will auto-convert any Microsoft Word or RTF document into a PDF file. All other formats require that the site administrator also contribute a suitable PDF file for posting.

The technical infrastructure for ingesting materials is provided by the Berkeley Electronic Press, a commercial enterprise started by faculty at UC Berkeley. The infrastructure they developed is based on a core belief that there are many inefficiencies in the current scholarly communication environment and that scholars must lead the way in developing new venues for their publishing needs. Bepress is a commercial entity with a number of other clients besides the CDL. One special interest of the company is to develop new electronic journals in underserved and emerging disciplines. The CDL licenses bepress software and contracts with them for setup, training, technical support, and rollout services. In the past, bepress and CDL had a co-development agreement under which bepress software, previously only used for journals, was enhanced to support repositories as well.

The bepress software manages both the peer review process and the process of creating large repositories. ProQuest has recently licensed it to provide electronic submission capability to graduate schools for electronic theses and dissertations.

Although the UC Press eScholarship Editions are stored in XML and delivered to the user in HTML, the Repository currently does not support XML. It is hoped that eventually the CDL will offer tools to allow for posting other content in XML, but this is still in a very early planning phase. The UC Press books are an exception because the texts were purchased from netLibrary with XHTML encoding already in place. That encoding was converted by eScholarship staff to XML in accordance with TEI. Materials in the Repository have basic metadata that is output as Dublin Core in compliance with the Open Archives Initiative Protocol for Metadata Harvesting. However, any subject terms for describing papers are entirely up to the unit that has established that series. No effort is made to control descriptive vocabulary within or among collections.

Current best practices for persistent access, according to CDL staff, “are reflective of the current state of the industry: regular backups are performed as it is housed in the access system where it first gets ingested and in which it resides.”<sup>6</sup> Longer term plans for the development of preservation framework for Repository materials and other types of electronic information managed by CDL will be developed with the support of a recently received IMLS grant.

Thus the UC eScholarship Repository will not rely on “self-archiving” by scholars for either access or preservation. But it is Open Archives Initiative (OAI) compliant to promote discovery of the content. In point of fact, over 50 percent of the page referrals currently come from Google indexing.

## **NEXT STEPS FOR THE REPOSITORY**

Currently there is no citation linking among Repository papers and external resources. Hitchcock et al.<sup>6</sup> make a compelling case for the development of open citation linking and analysis of the impact of open access scholarly materials. The implementation of “impact factors” for open access content could be a major step forward in the acceptance of Repository publications as worthy of consideration for promotion and tenure decisions. Without documentation of the visibility and impact of publications in the Repository it will be difficult to make any substantive inroads into the current print publication system. It might not be difficult to develop reference linking for the Repository items since the citation for each paper is already included in the HTML “title page” for each paper.

Another important step will be the development of support for online journals. At this point in time eScholarship staff are considering the pros and cons of a requirement for a “wrapper” for online issues that would mimic the print environment for journals. Current thinking leads them to conclude that a traditional journal format will be easier for contributors to “brand” and publicize, thus making this alternative attractive. Also, it may be easier for contributors to count publication for promotion and tenure if the open access journals in which they appear are similar to print journals.

It remains to be seen whether mimicking print journals or documenting the impact of individual articles, or some combination of both strategies, will be persuasive to the majority of faculty in moving from traditional print scholarship to open access scholarship. Guedon makes the case that “with the help...of [evaluations] that do not rest on the prior reputation of a brand, but on the actual quality of each selected work, librarians hold the key to developing a total, global mapping of science.”<sup>7</sup> Such a vision is attractive but its realization will require several developments related to interoperability of a wide variety of open access archives. One of the most important areas yet to be sufficiently developed is that of metadata protocols in open access archives.

If science, or any other discipline, is to be mapped sufficiently for that mapping to serve as a substantive alternative to the current scholarly communication system, some form of detailed indexing based on sufficiently granular metadata must occur. Dublin Core elements are not rich enough to provide sufficient searching across repositories for specialized resources or across disciplines that use similar terminology in disparate ways. Although vocabularies can be added by content providers it will probably take implementation of automated subject analysis and data mining based on full-text to overcome the laborious nature of in-depth subject analysis. The promise inherent in building a critical mass of open access scholarship will depend as much or more on the ability to do fast keyword searching across repositories, to automatically expand searches to related categories, and to rank results through robust relevancy algorithms than on mere exposure of Dublin Core elements to web indexers. Perhaps the experiment that the Research Libraries Group (RLG) is undertaking to mine their catalog through the use of Recommind Inc.'s MindServer<sup>TM</sup> technology<sup>8</sup> for automatic subject analysis will provide one answer that could prove useful for UC's eScholarshipRepository.

## CONCLUSIONS

The UC eScholarship initiative is focused on:

- Facilitating the expressed mutual interests of the University, its faculty, and the broader scholarly community;
- Leveraging the formidable capabilities and strengths of the University of California in order to provide effective national leadership in this area; and
- Supporting and extending the experimental reconfigurations of the components of scholarly communication by communities of scholars themselves.<sup>9</sup>

The eScholarship Repository is a vital element in accomplishing these goals. It is unique in that it has amassed high quality content in a very short time from exceptionally busy scholars. It has a flexible infrastructure that can accommodate both peer-reviewed and non-peer-reviewed content in a number of formats. Most importantly it is intended to be a long-term collaborative enterprise among librarians and scholars on all the University of California campuses. As Stephen Pinfield suggests, “this is a pivotal stage in the development of institutional open archives. We are moving from a position where awareness of the issues surrounding self-archiving was restricted to a relatively small

number of enthusiasts to a position where it is entering the consciousness of a large number of practice-based information professionals and some faculty.”<sup>10</sup> At this critical juncture in history, the eScholarship Repository has already proven its relevance to more than a few scholars. It holds great promise to leverage its current advantages into an even more robust example of open access scholarship in the very near future.

N.B. The opinions expressed in this paper are solely those of the author and do not imply agreement or endorsement of CDL staff.

## NOTES

1. Lynch, Clifford A. “Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age,” *portal: Libraries and the Academy* 3(2): 327-336. (2003)
2. <http://www.arl.org/sparc/core/index.asp?page=a0>
3. <http://mit.edu/mitworld/content/libraries/scdw.html#>
4. <http://creativecommons.org/>
5. <http://escholarship.cdlib.org/history.html>
6. <http://repositories.cdlib.org/escholarship/>
7. <http://www.bepress.com/>
8. <http://www.dlib.org/dlib/october02/hitchcock/10hitchcock.html>
9. <http://www.arl.org/arl/proceedings/138/guedon.html#v>
10. <http://www.rlg.org/redlightgreen/mining.html>
11. <http://www.slp.ucop.edu/programs/programsschol.htm>
12. <http://www.dlib.org/dlib/march03/pinfield/03pinfield.html>