

# Digital Letters

Spring 2004

Issue Number Five

## Letter from the Editor

Hello and welcome to the fifth issue of *Digital Letters*. This issue addresses two topics I hope you'll find of interest—open access and Internet portals. Shinjoung Yeo gives us insight into the reasons why open access is taking center stage in information communities, who the stakeholders are, and the role librarians can play. Internet portals, not unlike our own SAGE portal, attempt to create one-stop shopping areas for users to go to identify the most pertinent, high quality resources related to a topic. The Librarians' Index to the Internet (LII) is an example of such a portal geared towards public libraries and the patrons they serve. I took advantage of Karen Schneider's presence at the recent UCSD Privacy Forum and spent an hour talking with her about the LII service she currently manages. And to round things off, Dawn Talbot gives us a flavor of the Spring DLF Forum held this past April in New Orleans.

Laissez les bons temps roulez! ~ Trish Rose

### Librarians' Index to the Internet: A Conversation with Karen Schneider

According to the LII website, the mission of the Librarians' Index to the Internet is to "provide a well-organized point of access for reliable, trustworthy, librarian-selected Internet resources, serving California, the nation, and the world." More specifically, LII is a "searchable, annotated subject directory of more than 14,000 Internet resources selected and evaluated by librarians for their usefulness to users of public libraries. LII is used by both librarians and the general public as a reliable and efficient guide to Internet resources."

Trish Rose: So Karen, *who contributes to LII (i.e. who chooses the websites that are indexed?)*

Karen Schneider: Fifty percent of the content comes from volunteers, mostly public librarians from CA and the state of Washington, and the rest is staff contributed. While volunteers are not paid by LII they are paid by their institutions. Those libraries have agreed to allow their staff to spend a certain percentage of their paid time on this project because they recognize its value.

(Continued on page 2)

### The Open Access Movement: Knowledge as a commodity or a public good?

~Shinjoung Yeo, Reference Librarian, Social Sciences and Humanities Library

Earlier this year, *Outsell*, a research and advisory firm that focuses on the information content industry, announced their 2004 predictions. Their 6<sup>th</sup> prediction was, "The open access movement in scholarly and scientific publications will gain legitimacy" (<http://www.outsellinc.com/outlook/>).

The open access movement is an effort to provide free online access to scholarly literature including peer-reviewed articles, pre-prints and working papers. The premise of open access challenges the current subscription based system of scholarly communication that has become unsustainable due to the unfair pricing practices of an increasingly monopolistic publishing industry.

Many people quote the 2001 ARL study (<http://www.arl.org/news/tr/218/costimpact.html>) regarding skyrocketing journal costs as the impetus of the growth of the open access movement. However, those who are committed to open access models would argue that it is not primarily about costs, but about control over the scholarly communication process by the concerned players themselves – scholars and their communities, libraries, and academic institutions. Open access proponents ultimately consider knowledge as a public good and attempt to challenge the common view that knowledge is an economic commodity that can be bought and sold. In other words, this is not just about changing unfair pricing practices of scholarly publications, but transforming the entire academic culture – requiring scholars' willingness to publish their articles in open access journals and exploring new ways of communication as well as academic departments and disciplines' recognition of open access publications in the promotion and tenure process.

This cultural shift is going to take more than a few well intentioned academics and librarians. It necessitates a strong institutional commitment in implementing open access through allocating funds, implementing a system for institutional archives, and shifting policy to legitimize open access publications.

In recent years, more and more open access initiatives and

(Continued on page 3)

**LII** *(continued from page 1)*

TR: *How do libraries use LII?*

KS: Primarily, our most popular service is the current sites notifications that we send every week. This information goes out via email, RSS and is posted on the site. Some institutions may only visit the LII site once or twice a month because their primary updates are via email. We have 17,000 email subscribers and approximately 2,000 RSS readers. Yet, the portal is still heavily used and usage has gone up 25% every year.

TR: *And it began in 1991?*

KS: Yes, approximately 1991. Carole Leita, the founder, originally maintained LII as a gopher site.

TR: *It then moved to the Berkeley Public Library?*

KS: No, actually it stayed a volunteer project until 1997 when it was then funded by LSTA. Eventually it was called The Berkeley Public Library's Index to the Internet because Carole worked at that library and operated it from there but it was essentially a one person operation during that time. Once funding was secured she was able to hire some editors.



TR: *So the current staffing includes you fulltime and how many part time staff?*

KS: We have the equivalent of about 4.5 FTEs.

TR: *And how many volunteer contributors?*

KS: We have about 100 active volunteers who contribute at least a few sites every year. We've probably had up to 300 different people who have contributed to LII over the years.

TR: *So can you tell me about the technology behind LII – What have you been using and what are you transitioning too since you mentioned the architecture is changing?*

KS: The original software, written in PERL, was designed before content management systems existed. We're migrating to a modified version of Community Server (Community Server is open source server software for hosting weblogs written in Python). This software powers XML.com and other O'Reilly sites. It's being modified for us by the Community Server programmers for use in LII. It will be database driven so we can easily update our topic and browsing structure instead of having to hand-create static html pages. We will also be moving from natural language terms to using LCSH and follow Infomine's model and make them available as drop down boxes.

TR: *How do you handle dead links?*

KS: We have an active weeding program, for every three sites we add we delete one. I'd say we spend at least 25 hours a week on it. We devote as much time weeding our database as we do growing our database because we're only as good as our last search. People expect high-quality, current sites. The Community Server software will help with the maintenance of this. Westin, a non-profit educational company that we're working with, is helping us with iterative design in developing a maintenance workflow. This will make our portal unique and we hope will be a model for maintaining Internet portals. I'm on a subcommittee with LC that recently put out a report for bibliographic control of the web environment.

TR: *What sort of a niche do you see LII filling and are there are other services like it for public libraries?*

KS: We're unique in providing an announcement service for high-quality, newly discovered websites for librarians and general users. We focus on the life-long user community. It's an audience we know well as opposed to say the K-12 community. Our California focus is a nod to our funding community but also makes it unique. How do we compare with similar services? Internet Public Library (<http://www.ipl.org/>) doesn't have the quality control. Michigan Electronic Library (<http://mel.lib.mi.us/index.jsp>) doesn't have the announcement service. The Internet Scout project (<http://scout.cs.wisc.edu/>) and Infomine (<http://infomine.ucr.edu/>) both have an academic focus. The other thing we've done since Sept 11<sup>th</sup> is build and distribute current theme collections. Themes could include: war, the Patriot Act, Olympics, elections. We did a project with CDL where we identified, organized, and annotated over 400 digital library collections in the UC system. We exported that information to their database which they then display in their CDL portal. It provided them and us with desirable content.

*(Continued on page 4)*

## Open access movement

*(Continued from page 1)*

institutional digital repositories have been created as a strategy to promote open access. For instance, University of California eScholarship Repository (<http://repositories.cdlib.org/escholarship/>), MIT's DSpace (<http://dspace.org/>), Directory of Open Access Journals (<http://www.doaj.org/home>), Public Library of Science (<http://www.publiclibraryofscience.org/>) etc. show that institutions and scholarly communities are beginning to endorse alternative models of scholarly communications.

For libraries, the open access movement is a great opportunity to reposition ourselves within academia through using our traditional expertise -- knowledge collection, organization, dissemination, and preservation. Instead of passively reacting to the current market-oriented system, libraries are uniquely situated to be actively involved in creating and participating in a new knowledge infrastructure that will remove the barriers of current copyright law, licensing issues, and pricing models to scholarly communication brought on by digital technologies.

Converting the current subscription model to open access models is a complex and daunting task. There are still a lot of questions yet to be answered and we still do not know how the future will unfold. Only by exploring, experimenting, and pursuing alternatives, can we regain what has been lost --- control over universal access to knowledge.

**For more information on the open access movement, please visit the following web sites:**

**UCSD Biomedical and Medical Center Libraries**  
<http://scilib.ucsd.edu/bml/open/index.html>

**Lynch, Clifford A. (2003). Institutional Repositories: Essentials Infrastructure for Scholarship in the Digital Age**  
<http://www.arl.org/newsitr/226/ir.html>

**Issues in Scholarly Communication**[http://www.arl.org/scomm/open\\_access/](http://www.arl.org/scomm/open_access/)

**Open Access Now** <http://www.biomedcentral.com/openaccess/>

**Open Archives Initiative** <http://www.openarchives.org/>

**Arms, William, Y. (2003). Economic Models for Open Access Publishing**

<http://www.cs.cornell.edu/wya/papers/iMP-2000.html>

## Reports from the Field

### DLF Spring Forum

The Spring Forum for the Digital Library Federation was held in New Orleans, April 19 - 21. It was the largest attendance to date with about 160 participants—mostly programmers, digital project managers and directors of digital library programs. There were 13 attendees from Europe, Australia and Canada. Forum papers have been posted on the DLF site at <http://www.diglib.org/forums/Spring2004/springforum04abs.htm>. Dawn Talbot discusses highlights from the sessions she attended below:



### *Session 2: Virtual Collections*

A paper from the University of Michigan on "Digital Asset Management Systems (DAMS)" talked about commercial DAMS software they are working with to better manage digital content from campus communities. This is still in a pilot stage and they are encountering significant problems with interoperability of heterogeneous metadata that faculty have used and the lack of a preservation component. Howard Batchelor and Curtis Fornadley from UCLA spoke about their efforts to build a user centric digital library with a focus on the services that can be provided. Currently the collections they are working with each have their own siloed database. UCLA's focus has been working within on-line learning environments to integrate digital collections into applications for teaching and research. The final paper in this session was given by Thorny Staples from the University of Virginia (UVA) and provided a more in depth look at the architecture of the Fedora system developed at Cornell and being used at UVA to manage digital collections.

### *Session 4: Open Archives Initiative (OAI)*

Kat Hagedorn opened this session with a discussion of the University of Michigan (UM) experience as an OAI service provider. UM began development of the OAI repository service known as OAIster in June 2001 and it went public in June 2002. Today there are over 500 active data providers to that repository. Kat talked about the lack of any networking opportunity between data providers and service providers and the issues Michigan has encountered such as the lack of standards and the dearth of knowledge about content collections that would make the service providers job easier. Kat and her colleagues presented these issues at the recent OAI3 meeting in Geneva and as a result there is an interest in developing criteria for repositories, creating a registry of data providers where these issues can be addressed and developing crosswalks for metadata formats. Tom Habling from UIUC then spoke about an experimental registry of data providers they are developing that will be human readable and also machine parsable. Jeff Young from OCLC spoke about OCLC's ERRoL initiative - a Web browsable suite of services that makes using OAI repositories easier to use without data providers having to go to any extra effort. ERRoL used the UIUC experimental registry. More about ERRoL can be found at <http://www.oclc.org/research/researchworks/errro/default.htm>

### *Session 5: Metasearching*

The first two speakers in this session spoke about their efforts to provide metasearch functionality at their institutions - the first from Cornell using Endeavor's Encompass product and the next from N. C. State using a locally developed system. Both described the problems they encountered and several iterations they have tried in an effort to make something that is meaningful and useful to end users. The third speaker was Roy Tennant who spoke about the lessons CDL has learned in attempting to provide a metasearch capability beginning with UCSD's own Database Advisor and now CDL's SearchLight. Roy's list of metasearch issues and challenges is worth reviewing.

*(continued on page 4)*

**LII** *(Continued from page 2)*

TR: *How many new websites are listed in a weekly email announcement?*

KS: Via email we send out 30 newly discovered sites every week and we also remind people of sites we've already linked to on the portal that are relevant to a current topic (e.g. Memorial Day). We also post another 10-20 sites to the portal every week under the subject "more new sites" These are different from the 30 sent via email in that they may be of interest but perhaps not as high quality. Our users have asked us to include the "more new sites" information in the email so we'll be doing that in the future.

TR: *What do you see as the future of LII? Is it financially sustainable?*

KS: At least for the next couple of years the State of California has indicated it strongly supports this service. We've tried to pursue other revenue streams with varying levels of success. For example, we're approaching other states to build specialized portals like the one we did with Washington state. The Washington Project (wa.lii.org) is the Washington state version of LII. It uses the same database as LII and allows them to piggyback on our content but also add their own. WA librarians do the content and we provide the training. The key to LII is gradual growth. I think it will be around for a long time because it provides a quality service and fills a need.

**Reports from the Field** *(continued from page 3)****Birds of a Feather Session***

Several BDF sessions were held as the last session of the second day and are designed to be informal gatherings of attendees who share an interest in a topic and want to engage in a question and answer session rather than a formal presentation. I was primarily interested in the session on ARTstor because I was curious as to the audience for this new service that will be public July 1<sup>st</sup> and what type of questions the audience would have for James Schulman. There was a relatively small group – about 25 – and several of the CDL folks made up that number. Since the price model for ARTstor had just been announced there were some questions about the cost but more were concerned with the interoperability of the ArtStor platform with other presentation tools such as learning management systems, Luna Insight, and PowerPoint. James explained that restrictions placed on how the images could be used were in many cases constrained by the licensing agreements that ARTstor had entered into with the content providers. Designed as a repository of images for non commercial educational use, ARTstor expects to test the IP issues commensurate with providing access to a large number of images, many which are created by copy stand capture from copyrighted works.

***Session 9: Digital Images***

Laine Farley from CDL and Henry Pisciotta from Pennsylvania State spoke about their respective digital image projects, the Image Server Demonstrator Project (ISDP) and the Visual Image User Study (VIUS). Both projects aimed to assess user needs for the delivery of digital still images. Their findings, which contained not only an expected number of similarities but also a unexpected number of differences, can be viewed at: [http://library.ucsd.edu/~dawn/dlf\\_handout.rtf](http://library.ucsd.edu/~dawn/dlf_handout.rtf).

***Session 11: Preservation Repositories***

This session included a very interesting presentation by Jean Godby of OCLC on their Repository of Metadata Crosswalks that is designed to share previously created crosswalks across institutions. This repository utilizes METS in a most innovative way by modeling each crosswalk as a complex digital object within the METS encoding. In addition to being both human and machine searchable, services such as translations, queries, interactions with the data, and documentation can be built on top. The final paper in this session was jointly presented by David Greenbaum, Director of the Interactive University Project, UCB and Leslie Johnson, of UVa on digital repository interoperability. As digital repositories are being developed little attention is being given as to how to search across repositories and learning systems. For this reason, Mellon and the DLF helped fund the creation of a committee made up people from the digital library, learning systems, and publishing communities. This group will soon be issuing a checklist report for repository developers and the learning applications intended to work with them. It will focus on the flow of information from a repository to a user.

DLF's Fall Forum will be held in Baltimore October 25-27, 2004 and the Spring 2005 Forum will be here in San Diego, April 13-15 providing an opportunity for a limited number of people to attend.