

Digital Letters

Summer 2003

Issue Number Three

Letter from the Editor

Hello and welcome to the third issue of *Digital Letters*. This is the first issue of digital letters for me, Trish Rose, as its editor. As a brief introduction I have been working as a librarian here at UCSD since July of 2002. My job involves working on the Mellon-funded Union Catalog of Art Images (UCAI) project as a metadata analyst and database designer. With Steve Lawson's departure from UCSD, I volunteered to take over as editor of this publication which Steve so creatively brought to life.

My vision for this publication is not all that different from Steve's. While I will be reporting on issues being addressed by the Digital Library Program Working Group (DLPWG) at UCSD, topics in *Digital Letters* will not be limited by the extent of their activities. I plan to use *Digital Letters* primarily as a vehicle for sharing news and ideas, as well as exploring issues related to the creation, acquisition, management, delivery and preservation of digital resources (anything that can be measured in bytes) particularly at UCSD but also on a national and international level. I'm also hoping to demystify the concept of a digital library somewhat so that it becomes more relevant to our daily work and not just thought of as an experimental part of the library. Personally, I'll look forward to the day when the terms "digital" and "library" sound redundant and that the concept of library naturally includes digital and print without the need for distinctions.

I plan to continue Steve's use of interviews with folks at UCSD working on digital library issues and feature projects that support the library's mission through the use of technology. In addition, future issues of *Digital Letters* may contain subject bibliographies and book reviews by staff about publications they feel would make for insightful reads. I will continue Steve's Buzzword Bingo section (starting with this issue) because I like the idea of trying to pin down many of the terms and acronyms that get thrown around and used interchangeably in the digital library world.

I welcome any feedback, whether positive or negative, about this publication and ideas for future issues. Feel free to email me: trose@ucsd.edu or even pop by my office which is on the 1st Floor of the Geisel Library, Room 1046, just behind the music periodicals area. I look forward to meeting all of you eventually.

— Trish Rose

Reports from the Field

From May 14-16, 2003, Dawn Talbot and Linda Barnhart attended the Spring Forum of the Digital Library Federation in New York City. The Forum convenes twice a year and includes digital library practitioners from 31 member institutions, including the California Digital Library (CDL). UCSD staff attend as CDL members. This particular Forum was the largest ever held, with 180 people in attendance. An official summary of activities, and links to some of the presentations, may be found at the DLF Web site: <http://www.diglib.org/forums/spring2003/>.

The meeting began with an extraordinary keynote speaker, James Boyle, the William Neal Reynolds Professor of Law at the Duke University Law School, author of *Shamans, Software and Spleens: Law and the Construction of the Information Society*, and a founder and board member of *Creative Commons* along with Lawrence Lessig. Professor Boyle writes widely on issues of intellectual property, internet regulation and legal theory. He gave an intellectually stimulating, lively, and memorable talk on copyright, focusing on the "invisible losses" resulting from tighter copyright restrictions. The three key losses he described were the losses from the inefficient system for tracking and managing copyright information, the losses from failed sharing, and the losses from the misunderstanding of the concept of fair use. Boyle believes that libraries have a unique role to play and urged libraries to aggressively persist with asserting fair use.

Among the many concurrent breakout sessions, were several notable presentations. Roy Tennant and MacKenzie Smith were part of a panel that discussed library support for scholarly publishing. Roy used his time to talk about the decision at CDL eScholarship to go with proprietary software (bepress) rather than develop an open source solution such as the DSpace implementation at MIT. John Willinsky, a faculty member at the U of British Columbia spoke about the initiative he leads – the Public Knowledge Project, an open source publishing system designed for faculty. MacKenzie Smith started her remarks with a statement that journal publishing should not be construed as the whole story and that peer review should be re-examined and not taken as a given. She noted that DSpace is not a publishing system – it is an institutional repository and as such assumes that "publishing" has already taken place.

(Continued on page 3)

What and Where is diglet?

Some of you may be wondering if diglet is still in existence. Others may be wondering what the heck is a diglet anyway? Is it kin to a Chiclet? I'll let James R. Jacobs, primary contributor to diglet tell you all about it....

diglet is a weblog. Ok, all well and good you say, but what's a weblog? According to the Wikipedia, a weblog (or "blog") is...

"a website that tracks headlines and articles from other websites. They are frequently maintained by volunteers and are typically devoted to a specific audience or topic. Web logs are useful for web-surfers because they often collect numerous web sites with interesting content in an easy to use and constantly updated format."

Interest in blogs is growing by leaps and bounds, and there are many avenues where one can find in-depth information on them. In the bibliography below, I have picked out just a few articles on blogs and blogging that I think are especially relevant. It is by no means exhaustive.

So, back to the original question, "what's diglet?" diglet is a blog devoted to digital libraries. It can be found at <http://scilib.ucsd.edu/diglet/>. It is the frequently-updated online persona of the UCSD Digital Library Program Working Group's quarterly newsletter *Digital Letters*. diglet posts contain links to digital library projects and organizations as well as to the ever-growing issues surrounding them -- digital rights management, copyright, and emerging technologies like xml, etc.

diglet's front page will have the most recent posts. Previous posts are housed in the diglet archives. Each post contains a url, as well as a brief bit of text that will attempt to give some context, however skewed by my interests and opinions (but that's the fun of blogging!). I am experimenting with other blogging tools to make diglet more responsive and interactive ("sit, diglet, sit!"). In the future, diglet will include an RSS feed, automatically email interested people of recent updates, provide searchability, and allow readers to respond to posts. Please contact jrjacobs@ucsd.edu with kudos, suggestions, and/or harangues.

For more information about blogs I suggest:

- Carver, Blake (Winter 2003). Is it time to get blogging? *NetConnect*, 30 - 32.
- Embrey, Theresa Ross (December, 2002). You blog, we blog: a guide to how teacher-librarians can use weblogs to build communication and research skills. *Teacher Librarian* 30(2), p. 7 - 9.
- Fichter, Darlene (Jan-Feb, 2003). Blogging software for intranet applications: you can put your own creative juices to work thinking up ways to use Weblog software. *Online* 27(1), p.61 - 64.
- Fichter, Darlene (May, 2001). Blogging your life away. *Online* 25(3), p. 68 - 71.
- Jenkins, Henry (March 2002). Blog this (digital renaissance). *Technology Review* 105(2), p. 91.
- Thomsen, Elizabeth B. (2002). Blogging, anyone? *Collection Building* 21(2), p.76 - 77.

- Thanks for tuning in! James R Jacobs, Government Information Librarian

Want to become involved in the Digital Communications Group?

The Digital Communications Group is an umbrella organization for people involved in *Digital Letters*, diglet, and the Digital Dialogues series within the Digital Library Program at UCSD. Our mission is to educate, inform, and generate dialogue within the community at UCSD about digital library issues. We want to help bring these issues into the mainstream of library activities. This group works in collaboration with other library groups, such as LAUC-R&PD, in identifying issues of value to this community. As you know, *Digital Letters* and diglet communicate through print and online publications. Digital Dialogues is a series of workshops, presentations, and lectures by folks from UCSD and outside the campus.

Current members of the Digital Communications Group include: Elizabeth Cowles (coordinator for Digital Dialogues); James R. Jacobs (regular contributor and editor for diglet); and Trish Rose (editor for *Digital Letters* and coordinator for Digital Dialogues). We are looking for other enthusiastic folk to join the group. If you have an interest in following and sharing with others issues on the cutting edge of the application of technology in libraries please consider becoming involved in the Digital Communication Group. We welcome your participation and input! For more information contact dawn@ucsd.edu.

Reports from the Field

(Continued from page 1)

David Seaman, Director of the DLF, announced a new DLF initiative: the Distributed Open Digital Library DODL (pronounced "Doddle"), noting that this came from the DLF charter of 1995 that stated DLF would build a "distributed digital library". DODL proposes to make digital resources of major research libraries electronically accessible in a shared, unified collection of use for scholarship and teaching. David's PowerPoint can be viewed at http://www.diglib.org/forums/spring2003/Seaman-DLFToday_files/v3_document.htm.

Leslie Johnston from the University of Virginia gave a presentation on the FEDORA system that was balanced in terms of technical and user perspectives. While FEDORA has so far focused on text, it is now moving toward serving images as well. Leslie noted that Fedora, unlike DSpace, is a selected repository with all objects contained therein undergoing the same strict selection process that other collections undergo. For more details on Leslie's talk see: http://www.diglib.org/forums/spring2003/fedora_johnston_dlf.htm.

Gary Lawrence gave a thoughtful summary of the first analysis of UC's Collection Management Initiative (CMI) data. Jenn Riley and Michelle Dalmou from Indiana University presented their work on adding controlled vocabulary searching to a digital photograph collection. Forward progress was noted by the METS group (Metadata Encoding and Transmission Standard); METS will move towards becoming a NISO standard, and its proponents plan to hold a workshop in October in Washington, D.C. Several ARTstor staff members gave a breathless, whirlwind tour of the collection, technical, production, and metadata aspects of that product, which is expected to Beta test in the Fall.

Stephen Abrams, Harvard and MacKenzie Smith, MIT presented work to date on development of a Global Digital Format Registry that will aid digital library and archival communities with policy and processing decisions regarding ingest, storage, access, and preservation since these are usually format dependent. A steering committee has been formed to advise on development of the registry and CDL will be one of the 20 member institutions involved. More information about this project can be found at <http://hul.harvard.edu/formatregistry/>.

Other presentations by Deborah Holms-Wong of USC outlined the process they underwent to select and license a Digital Asset Management System (Documentum), and Herbert Von de Sompel talked about recent efforts to develop a file based approach to expose metadata via the OAI-PMH. Herbert's presentation is at: http://www.diglib.org/forums/spring2003/vandesompelOAI-SR_files/v3_document.htm.

— Linda Barnhart, Dawn Talbot

Buzzword Bingo

Yes, its back. Due to popular demand, this editor has decided to continue Buzzword Bingo as a regular section within *Digital Letters*. This issue's buzzwords are **OAIS** and **OAI**. Because these two acronyms are so similar and referred to frequently within the information community they can be easily confused with one another.

OAIS, or the Open Archival Information System, is a reference model that specifies the lifecycle of an information object from creation to management to dissemination in order to ensure the long-term preservation and access to that object. The **OAIS** is meant to be used only as a high-level conceptual model and not as a specification for an architectural implementation. Initially developed for NASA's space research community, it has since been adopted by some in the academic sector for managing digital repositories such as Harvard University's E-Journal Archive and MIT's DSpace.

OAI, aka the Open Archives Initiative, "develops and promotes interoperability standards that aim to facilitate the efficient dissemination of content". To facilitate this interoperability participants in this initiative have developed a metadata harvesting protocol, aka the OAI-PMH, for *data providers* to use to expose their metadata over a network. This content can then be harvested by a *service provider* who combines it with other content for the purpose of building value-added services. **OAI** initially developed out of the e-prints community which explains its loose interpretation of an archive as simply a "repository of scholarly papers". Those outside the community quickly realized its potential and it is now being used to provide access to a wide range of digital materials.

Though developed out of separate communities, both **OAIS** and **OAI** are being adopted within the library community for facilitating the sharing of information. Where they differ is in scope - the **OAI** has a much more limited focus because it does not address the long-term value and sustainability of information that the **OAIS** model attempts to.

For more information on **OAIS** I recommend

- http://www.ccsds.org/documents/so2002/spaceops02_p_t5_39.pdf
- <http://www.rlg.org/longterm/oais.html>

For **OAI** see

- <http://www.openarchives.org/documents/FAQ.html>

Will the trees continue to be sacrificed?

Unfortunately, *Digital Letters* will continue to be distributed primarily via paper. Like Steve, I too find myself reading primarily while having lunch, sitting at a coffeeshop, or riding home in my vanpool. Maybe I'm a throwback to another time but just so that others like me won't be left out in the cold on issues of "digital importance" I feel I must continue to provide *Digital Letters* in its byte-free form. That said, there will be a pdf version linked off of the DLPWG website (<http://gort.ucsd.edu/dlpwg/dletters/index.html>) and hopefully in the near future there will be an html version so you can quickly and easily follow the hot-linked urls to related content.

UPCOMING DIGITAL DIALOGUES

Monday, July 21st Seuss Room, Geisel Library

MacKenzie Smith

MacKenzie Smith, Associate Director of Technology for the MIT Libraries and project manager for DSpace, will be here on Monday July 21st to talk with us about DSpace - MIT's ambitious Web-based institutional repository. DSpace, an open source software platform is a specialized type of digital asset management or content management system designed to capture, preserve and communicate the intellectual output of MIT's faculty and researchers. The DSpace software is being tested by six other US institutions who will help test and improve the software. Altogether 3,400 organizations have downloaded the software since it was released in November. We have still to confirm the time but we are expecting her presentation will be in the Seuss room at around 1:30 p.m. Look for more details via email.

Monday, August 25th 2-3:30pm Seuss Room, Geisel Library, *refreshments will be served

Daniel Pitti—Challenges in Library Collecting of Original Digital Humanities Research Publications

Pitti's presentation will focus on some of the many intellectual, technical, and political challenges of library collecting of complex humanities research publications. In particular, the presentation will focus on selected projects created and maintained at the Institute for Advanced Technology in the Humanities (IATH) at the University of Virginia and a 3-year Mellon funded project to explore library collecting of these projects.

Daniel Pitti is Co-interim Director of the Institute for Advanced Technology in the Humanities (IATH) at the University of Virginia. His permanent position at IATH is Project Director. As Project Director, Pitti is responsible for project design in general, and Extensible Markup Language (XML) and object-relational database design and development in particular. Before coming to IATH in 1997, Pitti was Librarian for Advanced Technologies Projects at the University of California at Berkeley Library. Since 1993, Pitti has been the chief technical architect of Encoded Archival Description (EAD), an international standard for encoding library and archival finding based on SGML and XML. Most recently, Pitti worked with the EAD Working Group in the revision of EAD version 1.0. He is also working with an international group of archivists developing a prototype standard for representing description of individuals, families, and corporate bodies, Encoded Archival Context (EAC).

At Berkeley, Pitti assisted in establishing the California Heritage Collection and the UC-EAD Project (later the Online Archive of California). At IATH, among other projects, Pitti has been involved in the design and development of the following: the Walt Whitman Archive; Lives of Exemplary Women; Salem Witch Trials; Evolutionary Infrastructure: Boston's Back Bay Fens; Monuments and Dust: the Culture of Victorian London; Aquae Urbis Romae: the Waters of the City of Rome; the Salisbury Project; the William Blake Archive; the Thomas MacGreevy Archive; Uncle Toms Cabin; The Samantabadhra Archive; and The World of Dante. In addition, he has served as a consultant to the Dante Gabriel Rossetti Archive, the Piers Plowman Archive, and many other projects.

Friday, October 24th 1:30-2:30pm Seuss Room, Geisel Library

Amy Warner—Metadata, Content Management and Interoperability

Warner's talk will address the issues and methods associated with organizing and managing content where a central focus is on effective use of metadata. The kinds of metadata solutions will briefly be reviewed, as well as their role in effective content management. A major part of the talk will address interoperability issues associated with both structural (e.g., XML) and content (e.g., controlled vocabularies) metadata. It will be argued that interoperability at the structural level only solves part of the interoperability problem. Strategies and methods of achieving both structural and partial content metadata interoperability will be detailed.

For more than fifteen years, Amy Warner has practiced, taught and researched the organization of information. A former associate professor at the University of Michigan School of Information, Amy is an expert in bottom-up information architecture, controlled vocabularies, and metadata and thesaurus design. She applies these skills to create coherent, effective information environments on public Web sites or corporate intranets. As a consultant, Amy has worked with Fortune 500 companies, the U.S. government and several academic institutions to develop controlled vocabularies and site indices and evaluate the effectiveness of existing thesauri. Another of her specialties is the integration of thesauri with search engines and content management systems. She draws on her teaching background to provide on-site training in thesaurus construction and indexing methods.