DIGITAL LIBRARY FEDERATION

Final Schedule
DLF Fall Forum 2002

Pre-Forum

Monday, November 4

9:00-1:00: **Developers Forum.** *Taku/Chinook Room.*

10:00-1:00: **E-resources Management meeting (Tim Jewell, chair)** *Executive Boardroom.*

10:00-1:00: **E-metrics meeting (Denise Troll Covey, chair).** *Makani Room.*

Forum

Monday, November 4

1:00-2:00: Registration. *Pacific Ballroom foyer.*

2:00-3:30:

Welcome and Introduction: David Seaman, Digital Library Federation. *Pacific Ballroom.*

General Session 1: Bill Hill, Microsoft Research: "From Molecules to Bits: Reading in the 21st Century." *Pacific Ballroom.*

Although eBooks have been slow to take off, reading on the screen is fast gathering momentum and is about to go mainstream. Microsoft has focused major efforts on making onscreen reading more like reading paper. TabletPC will launch on November 7 as the first PC designed as a reading surface. Bill Hill will talk about his vision of the future of reading on the screen.

3:30-4:00: Break

4:00-5:30: **Breakout Session 1: eLearning.** *Windward Room.*

Denise Troll Covey, Carnegie Mellon University: "Filling the Gap Between Vendor and User Practice."

Results from ARL's LIBQUAL+ project indicate that personal control is a high priority for library users, but that users are not very satisfied with the ease of remote access to library resources. This gap between user preferences and the level of service they perceive their library providing is exacerbated by the gap between the way commercial vendors restrict access to library resources and the way users access those resources. Restricting access by institutional IP address creates problems for the increasing number of institution-affiliated users using computers that do not have an institution-affiliated IP address. The Council on Library and Information Resources recently sponsored a survey of how academic libraries are supporting remote access from non-institution-affiliated IP addresses, the problems and costs associated with this effort, and their level of satisfaction with the technology they've implemented. Preliminary results reveal that libraries are using proxy servers and (to a lesser extent) virtual private networks to fill the gap between vendor and user practice, but that these solutions can be expensive and unsatisfactory. Proxy servers appear to be more problematic than VPNs, but both technologies lower user satisfaction, service quality, and staff morale; cause delays in other projects; and have a detrimental impact on the allocation of resources. Moving to a new solution, however, will require time, money, proof of adoption by many vendors, and sufficient confidence that the new technology will be superior to the old and that the transition will be rapid and transparent to users.

Fred Beshears, University of California, Berkeley: "Learning Technology Application Infrastructure and Interoperability Standards."

Interoperability among system software components and online learning content is key to the successful implementation of a campus software architecture, and several learning-technology standards are being promoted to help schools achieve this objective. The purpose of the first half of this presentation is to give the library community a conceptual model of the different services educational technologies provide and how they can be integrated into an overall campus software architecture. The second half of the presentation will provide a basic understanding of the main learning technology standards that pertain to higher education, the key organizations

involved in developing the standards, and the objectives they seek to achieve. UC Berkeley has been an active participant in IMS, one of the major learning-technology standards-setting efforts. Fred Beshears served as the UC System's representative to the IMS technical board for two years starting in 1998, and he has been UC Berkeley's representative to IMS since the end of 1999. Fred is also Berkeley's liason with the Open Knowledge Initiative.

4:00-5:30: **Breakout Session 2: National Digital Information Infrastructure and Preservation Program.** *Leeward Room.*

Abby Smith, Council on Library and Information Resources: "LC report to Congress."

In December 2000, the United States Congress passed legislation establishing the National Digital Information Infrastructure and Preservation Program (NDIIPP). It charges the Librarian of Congress to lead a nationwide planning effort for the long-term preservation of digital content. The goal of the plan for digital preservation is to encourage shared responsibility for digital content and to seek national solutions for:

- the continuing collection, selection, and organization of the most historically significant cultural materials and of other important information resources, regardless of evolving formats
- the long-term storage, preservation, and authenticity of those collections
- persistent, rights-protected access for the public to the growing digital historical record and information resources of the American people.

The process began with a year-long, nationwide, fact-finding effort and initial planning. Next, with Congressional approval, NDIIPP will invest in a set of activities proposed under the plan that include practical applications and modeling of key components of the infrastructure; developing core capacities for the preservation network; building a digital preservation architecture; and conducting targeted basic research needed for the management of digital content and of the systems that support it. These investments will both leverage the knowledge gained by a range of preservation stakeholders and broaden their participation in network building.

Abby Smith will report on the results of the planning phase and the recommendations made to Congress this fall for implementation of the plan.

6:00-9:00: Reception. *Soundview Room.*

Tuesday, November 5

8:00-9:00: Continental Breakfast. *Pacific Ballroom foyer.*

9:00-10:30: **Breakout Session 3: Open Archives/Cultural Archives.** *Windward Room.*

Martin Halbert, Emory University, *et al*: "Major Findings from the Mellon Metadata Harvesting Initiative."

> This session contains presentations and discussion on the various aspects and institutional projects comprising this large and significant initiative. The Metadata Harvesting Initiative of the Mellon Foundation is a $1.5M effort to explore and test applications of the OAI PMH (for more background details see http://www.arl.org/newsltr/217/waters.html). The seven projects funded by the Mellon Foundation have made significant progress in advancing the understanding and use of the OAI PMH for harvesting and other associated scholarly services.
> Speakers will review the goals of the Mellon program, and discuss findings concerning core OAI technologies, such as best practices in creating OAI metadata providers and harvesting tools created during the projects. Speakers will describe services built upon OAI metadata harvesting infrastructures, such as portal features that add value to aggregated metadata, user interfaces to harvested metadata, and will discuss issues in accommodating different metadata formats, and future directions for research.

9:00-10:30: **Breakout Session 4: Integrated Cultural Resources.** *Leeward Room.*

Panel Discussion: Access to Integrated Cultural Resources. Moderator: Bob Wolven (Columbia).

> In the last few years many of the digitization projects in the community have focused on capturing and bringing together surrogates of cultural resources. These projects are at various stages of maturity and reflect various approaches to the range of challenges inherent in the aggregation of digitized materials. With this body of pooled collections, now may be a good time to reflect on the benefits of the various approaches. Some are distributed; others are centralized; while others use a combined approach. But in each

case, the goal is to improve access to collections that are of interest to researchers and educators.

This session will provide an opportunity for demonstrations and discussion on some of the persistent issues that continue to confront these effort, such as:

- Promoting Use -- What is being done to increase awareness of the resource, how can researchers access it, what is being done to enable use of the content with courseware or in classrooms and projects, and are other uses allowed?
- Rights Management -- What materials are included, how are they protected, and what agreements are in place between the project and the content owners and between the project and the users?
- Technical Integration -- Is the content centralized or distributed, how are metadata in a variety of formats and from a variety of sources made meaningful to users, and how are digital files that were created in a variety of ways presented to users?
- Sustainability -- What will happen when project funding is discontinued, how will the resource be kept alive and continue to grow, and how are the materials being preserved for the future?

10:30-11:00: Break

11:00-12:30: **Breakout Session 5: E-Scholarship.** *Windward Room.*

David Loy and Suzanne Samuel, California Digital Library: "Acquiring and Ingesting Digital Content: Methods of the CDL's eScholarship Program."

This team presentation by the eScholarship project coordinator and the eScholarship technical lead will discuss technical and programmatic issues involved in the acquisition, conversion, and dissemination of digital content. Three projects will be used as case studies: the 1500 University of California Press titles that eScholarship is converting from netLibrary XHTML files into XML ebooks; the eScholarship Repository, an institutionally based repository for pre-publication scholarship in all disciplines; and the Handbook of Visual Optics, a reference work that will appear in both digital and print forms from a collaboration among CDL, UC Press, and the Optical Society of America.

Programmatic topics will include challenges to 2 and 3-way partnerships (across libraries, university presses, and scientific societies); collections issues of university-specific access vs. global access; and institutional vs. subject or faculty-based repository arrangements. Technical topics will

include the conversion of incoming objects to METS; automatic conversion and merging of metadata from MARC; a processing database to MODS and to DC; exposing results through OAI; tools for authoring and publishing; and strategies for searching XML-encoded texts.

Thornton Staples, University of Virginia: "Report on the 'Supporting Digital Scholarship' Project."

The Supporting Digital Scholarship project is a joint project between the Instititute for Advanced Technology in the Humanities and the University of Virginia Library to investigate the implications of collecting digital scholarly projects into a digital library. This project, now in its third and final year, was funded by a grant from The Andrew W. Mellon Foundation. Our experience at the University of Virginia has been that scholars, given the resources and the support needed, are creating digital projects that are more like virtual museum exhibitions than like books. These projects usually include large collections of digital versions of primary resources with a network of complex interrelationships interwoven with original scholarly commentary. This report will discuss the technical problems that we have encountered in collecting the Rossetti Archive (
http://jefferson.village.virginia.edu/rossetti/) and the Salisbury Project (http://www.iath.virginia.edu/salisbury/) into the Library's Feodora repository, and policy implications for the Library in collecting such projects.

11:00-12:30: **Breakout Session 6: Metadata Object Description Schema (MODS).** *Leeward Room.*

Rebecca Guenther, Library of Congress; Kirk Hastings, University of California; Mary Alice Ball, University of Chicago Press: "Metadata Object Description Schema (MODS): An overview of its uses, features, and current experimentation."

This session will consist of several presentations on MODS, a new bibliographic element set based on MARC and represented as an XML schema. MODS has been developed by the Library of Congress with interested parties as a rich element set that is a subset of MARC, using XML with language based tags. Potential uses include, among others, use as a specified format for Z39.50 Next Generation, as an extension schema to METS, to represent metadata for harvesting from databases, and as a simplified record in XML that is compatible with existing library data. Part I. Development of DC-Lib Application Profile: as background to understanding the development of MODS, there will be an introduction to the reasons for developing the Dublin Core-Library Application Profile,

what it attempts to accomplish, and its current status. Implementation for this profile may be touched on briefly, and simple Dublin Core will be addressed as discussed below.

Part II. Introduction to MODS An overview of the reasons for developing MODS, an introduction to the element set, and a summary of potential uses. The simple Dublin Core element set will be contrasted to MODS in terms of the details in the element sets, and some reasons for using MODS rather than simple Dublin Core. (Speaker: Rebecca Guenther)

Part III. Descriptions of MODS experimentation, including:

- University of California project using METS with MODS as descriptive metadata to support a search/browse interface (Speaker: Kirk Hastings)
- University of Chicago Press project using MODS to enhance electronic document searchability and distribution (Speaker: Mary Alice Ball)
- Web archiving project at LC using MODS for creation of minimal metadata to support search and retrieval of archived Web sites. (Speaker: Rebecca Guenther)

12:30-1:15: Lunch. *Soundview Room.*

1:15-2:00: **General Session 2: David Levy, The Information School, University of Washington: "To Grow in Wisdom: A Response to the Crisis in Libraries and Universities."** *Soundview Room.*

What role can and should libraries play in the future of the university? Much of the focus today has been on scholarly publishing and information organization and delivery. The speaker will argue that publication and information (or knowledge) production are only part of the story, and that the library, especially at this time of growing "virtuality," carries and symbolizes other values and modes of being that are essential for the future of education.

2:00-3:30: **Breakout Session 7: MARCXML and MIX.** *Windward Room.*

Jerome McDonough (NYU), Morgan Cundiff (Library of Congress), Rebecca Guenther (Library of Congress), and Robin Wendler (Harvard): "MARCXML (descriptive metadata) and MIX (technical metadata for still images)."

The METS (Metadata Encoding and Transmission Standard) Schema Version 1.1 is now reasonably stable. One of the issues the community of METS implementers faces is which (if any) extension schema to use for descriptive and administrative metadata to complete a METS information

package. This panel presentation chaired by Jerome McDonough will provide information about several schemas that can be used as METS extension schemas including MARCXML and MODS (for descriptive metadata), MIX (technical metadata for digital images), and other schemas for text, audio, video, and rights.

2:00-3:30: **Breakout Session 8: OAI and Unicode.** *Leeward Room.*

Caroline Arms, Library of Congress: "Implementing the OAI Protocol for Metadata Harvesting at the Library of Congress."

The Library of Congress has participated in the development of the Open Archives Initiative Protocol for Metadata Harvesting since the first meeting in Santa Fe. LC joined the technical discussions and tests before the release of the first and second public versions. Released this June, version 2 includes a recommendation that harvesting of MARC records be based on a new, "slim" XML Schema for MARC21 maintained by LC's Network Development and MARC Standards Office (NDMSO). Over 100,000 metadata records for items digitized for American Memory are now available for harvesting. Items represented include books, maps, photographs, early movies, sheet music, and printed ephemera. Records for more content that LC can share freely will be added steadily. Several organizations building union resources of cultural heritage materials, including DLF members funded by the Mellon Foundation, have harvested the records. The records are available in MARC and unqualified Dublin Core, with MODS to be available soon. Metadata crosswalks and character set mappings developed by NDMSO made it easy to implement the straightforward protocol as an add-on to American Memory. The requirements for supporting the protocol have induced discipline and consistency with long-term local benefits and provided test cases for handling special characters in UNICODE. Two specialist gateways in the planning stages, for movies and sheet music, expect to harvest records for particular collections. Other collaborations in which LC is involved also plan to make use of the protocol. LC has found it valuable to participate in the experimental phase of this contribution to the toolkit for collaboration.

John Walsh, Indiana University: "Unicode Ate My Brain: The Trials and Tribulations of Implementing a Unicode-Dependent Digital Library Project."

The speaker will discuss the various challenges and problems raised in implementing a digital library project dependent upon Unicode data. During the past two years Indiana University's Digital Library Program has been

working on a project--funded by a grant from the Technological Innovation and Cooperation for Foreign Information Access (TICFIA) program of the U.S. Department of Education-- to digitize and make available on the Web a twenty-year run (1956-1975) of the *Letopis Zhurnal'nykh Statei*, a Russian/Soviet periodical index covering virtually all disciplines. The language of the index is Russian and uses the Cyrillic alphabet but includes other languages and scripts as well, including English, French, German, and Greek. This combination of languages and scripts necessitates the use of Unicode in the digitized version of the index. A Web-based digital library project that is dependent upon Unicode poses a number of challenges and requirements, including full Unicode support in the programming languages and other technologies used to implement the project. There are also special considerations and problems regarding fonts, browsers, display of right-to-left languages, and user input. Using the *Letopis Zhurnal'nykh Statei* project and other Indiana University projects as examples, the speaker will discuss these and other challenges presented in working with Unicode data and some of the solutions devised to address them.

3:30-4:00: Break

4:00: **Birds of a Feather Sessions**

a) **METS implementation. Jerry McDonough, chair.** *Soundview Room.*

Digital libraries are beginning to employ the METS digital library object schema for a variety of purposes. This informal session will allow those working on implementing software systems which use METS to present the work and ask for feedback and guidance on future development. Possible topics for discussion include:

- Developing local extension schema
- METS object viewing software
- METS and XSLT
- Using RDBMS systems for storing METS objects/metadata.

b) **Libraries as publishers and/or collaborators in publishing. Maria Bonn, chair.** *Windward Room.*

This will be an information-sharing session for insitutions that are developing or thinking about developing electronic publishing programs. Time will be reserved to discuss tools for electronic publication, particularly peer review and manuscript management systems.

c) Developing a model and standards for agreements to establish and document the intellectual property rights associated with digital projects. Joyce L. Ogburn, chair. *Chinook Room.*

At the May 2002 DLF Spring Forum, a Birds of a Feather session addressed the need for standardizing agreements to establish and document the intellectual property rights associated with digital projects. The initial proposal stated that we should cooperate in drafting model licenses that support the rights for libraries to create or host digital content in the same way that we banded together to understand terms and build model licenses for licensing content from commercial suppliers. The attendees agreed that the session was an invaluable start in exploring the issues and identifying some concrete actions. This session will build on the previous one and go further in outlining steps for developing a model and standards. It will also explore the possibility of linking this effort with those focused on management systems and metadata for electronic resources.

d) Humanities computing centers in libraries. Jennifer Vinopal, chair. *Taku Room.*

More and more academic libraries, either by themselves or collaboratively, are creating services to encourage and support humanists' use of computing technology for scholarly purposes. These services may be offered within a dedicated Humanities Computing Center, or may be an outgrowth of pre-existing facilities (computer lab, e-text center, etc.). In this informal discussion, participants will share experiences and explore issues related to libraries' provision of humanities computing services in such facilities. Possible discussion topics include:

- staffing and staff training
- collaboration with other departments, units, or academic divisions
- marketing your service and building your clientele
- budgeting time and money
- growing the service along with user demand
- project support (staffing, funding, etc.)

One possible outcome of this session may be to create a forum, either informal or formal, in which participants can continue sharing ideas after the conference.

e) Multilingual data and the Digital Library. Eileen Llona, chair. *Makani Room.*

Much of the information that academic libraries hold is multilingual in format. Making these data available to digital library systems involves several considerations, and often several systems. Digitizing text involves OCR if the text is to be searchable, which often raises language support issues for non-Roman languages. Storing digital data from scratch in non-Roman languages requires specialized software. Combining the richness of mutlilingual information into one system presents challenges for interface design and back-end systems. The University of Washington has been actively creating new digital resources involving non-Roman languages, and will lead a discussion of multilingual access in a digital library, including:

- Character encoding issues in an ILS
- Integration of multilingual data into thesauri development
- Character encoding and font issues for integrating multiple systems involving multiple languages.

Related topics are welcome. The goal is to share experiences and solutions, and discuss the potential for collaboration.

f) The Digital Library Federation Electronic Resource Management Initiative. Tim Jewell, chair. *Leeward Room.*

The DLF is sponsoring an initiative aimed at developing common specifications and tools for managing the license agreements, related administrative information, and internal processes associated with collections of licensed electronic resources. The project's goals are to describe the necessary architectures, establish lists of data elements and definitions, write and publish appropriate XML schemas, and identify and promote best practices and data standards. This session will describe the project and work plan, summarize a preconference meeting of the project's working group, and solicit comments on the project plan and local development efforts from interested DLF members.

Wednesday, November 6

8:00-9:00: Continental Breakfast. *Pacific Ballroom foyer.*

9:00-10:30: **Breakout Session 9: Music and Math.** *Windward Room.*

Jon Dunn and Mark Notess, Indiana University: "The Variations2 Digital Music Library System."

The Variations2 digital library system is being developed by Indiana University, with funding in part from a Digital Libraries Initiative Phase 2 grant from the National Science Foundation, to provide improved access to digital music collections, enable the development of applications for music learning and research, and support a program of digital library research in the areas of instruction, usability, and intellectual property rights. The basic goals of the Variations2 project were discussed in a presentation at the Fall 2000 DLF Forum. Since that time, much work has transpired, and version 1.0 of the Variations2 system was recently completed, offering integrated access to sound recordings and musical score images through a Java-based client application and back-end server, and utilizing a data model and metadata scheme based in part on IFLA's Functional Requirements for Bibliographic Records (FRBR) work to offer enhanced capabilities for finding and retrieving musical content.
In this presentation, Variations2 version 1.0 will be described and demonstrated, including discussion of the technical architecture, data/metadata model, and user interfaces for search and navigation. The relationship between the Variations2 data model representation and METS (Multimedia Encoding and Transmission Standard) will be discussed. In addition, the status of current work on multi-format synchronization, support for music notation file formats, and integration of digital music collections and services into instructional applications will be presented.

David Fielding, Cornell University: "D-Pubs, Project Euclid, and OAI Compliance."

The Digital Publishing System (D-PubS) is being developed at Cornell University for Project Euclid (http://ProjectEuclid.org), with support from The Andrew W. Mellon Foundation. Project Euclid strives to facilitate affordable scholarly communications in the fields of mathematics and statistics. D-PubS is the modular software architecture developed for Project Euclid. We have extended and generalized the original system developed for Project Euclid to support a broad variety of digital collections, including the Cornell University Library Technical Reports and Papers collection and the Historical Mathematics Book collection. D-PubS consists of a set of configurable services that may vary depending on the functionality desired for a specific collection. D-PubS services include a repository service, index and search service, user interface service, pay-per-view service, subscription

service, registry service, editorial workspace service, and a moderated submission service. A peer-review service is planned. In attempting to follow standards, D-PubS currently supports the Open Archives Initiative (OAI) metadata harvesting protocol and we are planning to adopt the design guidelines of the Open Archival Information System (OAIS) reference model. The speaker will give an overview of the D-PubS system and discuss our experiences implementing the open archives protocol along with our plans to develop and/or integrate advanced OAI services into the D-Pubs system.

9:00-10:30: **Breakout Session 10: "The Digital Library and the Classroom."** *Leeward Room.*

Oya Y. Rieger (Cornell), "Integrating Digital Libraries with Virtual Learning Environments."

One of the challenges we are facing is how to achieve the seamless integration of digital libraries in learning management environments. This presentation will focus on Cornell University Library's approach in building a service infrastructure and forging alliances with key stakeholders to ensure the incorporation of digital library resources and services in virtual learning environments.

As manifested by the increasing number of course Websites and online course modules, faculty are making significant use of information technologies in support of teaching, learning, and research. The goal of the presentation is to share strategies developed by the Cornell University Library (CUL) to support faculty initiatives in technology-rich environments. These programs are collaborative in nature and bring together several campus service providers to create a seamless support framework for technology-mediated learning. The presentation will highlight two tracks of the CUL's distributed learning strategy: the Unified Service Working Group and Digital Consulting and Production Services.

Co-managed by the CUL, the Unified Service Working Group is a campus-wide initiative to coordinate service access for faculty interested in using various distributed learning technologies. As distributed learning services proliferate, the faculty needs systematic assistance in identifying relevant resources and services in support of their specific projects.

The goal of the CUL's new Digital Consulting and Production Services (D-CAPS) is to offer a suite of digital asset production and management services to the campus community, including unit libraries, faculty, and administration. D-CAPS is comprised of associated services necessary to ensure cost-effective planning, creation, management, use, and preservation for digital collections. D-CAPS consults in the areas of digitization, copyright, metadata, technology, digital preservation, and digital asset

management.

As illustrated by these examples, integrating digital libraries in learning management environments requires the development of a collaborative service infrastructure. The presentation will focus on CUL's approaches in building this framework.

Cathy Marshall, Microsoft Research: Reading in the Digital Library: Experiences with E-books in Education

Digital libraries are delivering on the promise of anytime, anywhere access to educational materials. But how will students read these materials? Are e-books the answer? In this talk, the speaker will discuss implications of the results of several studies of reading, reading technologies, and the use of digital library materials both in and out of the classroom.

10:30-11:00: Break

11:00-12:30: **Breakout Session 11: Enhancing the User Experience.** *Windward Room.*

Peter Brantley (NYU), Access management: Shibboleth introduction; Scott Cantor (OSU), and Oliver Pesch (EBSCO): Shibboleth implementation issues

The Internet2-backed Shibboleth middleware initiative has progressed rapidly in the last six months. Initial production code is slated to be released in late October as part of the NMI Release2 package. Significant new vendor support has been announced, and exciting integration work is underway with leading content providers and learning management systems.

The initial session will review the status of Shibboleth software and the organic development of enhancement proposals and vendor interest in Shibboleth compatability. Following this introduction, Scott Cantor will technically review Shibboleth message flow and then talk about the implementation experience at his source site, Ohio State. On the target side, Oliver Pesch will discuss EBSCO's basic authentication set-up, their incorporation of Shibboleth and its advantages to the user community, and the issues of mapping accounts to institutions.

11:00-12:30: **Breakout Session 12: Registries and Archives.** *Leeward Room.*

Taylor Surface, OCLC: "Digital Registry"

The Digital Registry working group has completed a draft document providing guidelines for using MARC21 in the current OCLC cataloging system. The working group participants will review the guidelines and provide examples of usage on actual digital master project materials. The working group will also be reviewing project progress and soliciting broader participation in reviewing the guidelines through hands-on use of the Digital Registry.

Michele Kimpton, "The Internet Archive -- building a digital library of creative human works -- progress and prospects."

The Internet Archive has been working in collaboration with numerous partners to put together the an ambitious digital collection of works, including websites, movies, animation, TV, music, and books, as a tool for providing universal access to human knowledge. The vision of IA is to improve access to all these materials by removing geographic, physical, and monetary barriers wherever possible. Over the last year, IA has been working in partnership with several organizations to help them build, store, and access such collections by providing digitizing services, unlimited disk storage and bandwidth, and online collecting that might otherwise be out of reach. The speaker will provide an overview of current IA projects and collections. A more in-depth discussion will be given of a current project to collaborate with National Libraries to archive the Web.

12:30-1:00: **Closing remarks: David Seaman.** *Pacific Ballroom.*

Post-Forum:

Thursday, November 7

10:00 - 4:00: METS Board Meeting. Jerry McDonough, chair. *Makani Room.*