Breakout Session Group One: Thursday, June 25, 2009, 10:30-12:30				
(1) Embedded Metadata and Extraction *Churchill Room*	(2) Organizational Agreements for Collaboration *Balcony*	(3) Collections and Storage *Ballroom*	(4) Digital Preservation of Dynamic Content *Hunt Room*	(5) Negotiations and Incentives *Salon*

Breakout Session Group Two: Thursday, June 25, 2009, 1:30-3:30				
(6) Making Metadata Do Something Useful *Balcony*	(7) Building Community and Partnerships *Hunt Room*	(8) Tools *Ballroom*	(9) Access and Web Archiving *Churchill Room*	(10) Content Creators and Metadata *Salon*

Please see the following pages for detailed session topics

1. **<u>Breakout Session One: Embedded Metadata and Extraction</u>**
   Churchill Room

(A) *Metadata Extraction and Web Archives: Automating the Record Creation Process*
Presentation
The Library of Congress Web Archives (http://www.loc.gov/lcwa) interface provides item-level catalog record access to thousands of archived web sites. To streamline the record creation process, the Library uses automated techniques to build the records. Metadata such as keywords, abstract, document title and capture dates is extracted from the Web Archive data, and additional data is pulled from our URL nomination and permissions tool. The resulting initial record is incorporated into an XML template using the MODS schema to create metadata records that catalogers can review and enhance for use on the LCWA site. These provide access through subjects, names and titles to the web archives. This presentation will describe the extraction process and next steps that occur to create Web Archive MODS records.

Presenters:
Abbie Grotke, Library of Congress
Tracy Meehleib, Library of Congress

(B) *Using Embedded Photo Metadata to Enhance the Value of Image Collections*
Discussion Session
This session will discuss how value can be added to an image collection by embedding information directly into the file itself (standard TIFF and JPEG files) using the right metadata fields, standards, and tools. Having the right information within the image file can save you time later, ensuring that your images can be found by yourself, your staff, and your patrons, especially when sharing these images over a network or website. The importance of captions and keywords will be discussed, and the use of proper metadata standards and meta-friendly tools will ensure files can be easily accessed and read by others.

We'll share some tips and hints to help you over the hurdles for getting metadata into your images, quickly and painlessly. We will demonstrate how to do this with a couple of the more popular software tools including Photoshop, Lightroom, Photo Mechanic and Expression Media.

Presenter:
David Riecks, Stock Artists Alliance

2. **Breakout Session Two: Organizational Agreements for Collaboration**
    Balcony

(A) *Mutual Digital Preservation Contracts and Agreements*
Panel Discussion
Many digital preservation strategies to date have focused on mutual preservation arrangements, but a shared understanding of how to create formal written contracts and agreements for such collaborative efforts is still emerging. This panel will allow NDIIPP partners who have developed or are in the process of developing mutual digital preservation agreements to share documents (where possible), give brief presentations, and then engage in an extended discussion. The intended outcome is to identify the salient findings that are emerging on such agreements in the interest of developing a shared understanding of best practices.

Presenters:
Dwayne Buttler, MetaArchive Cooperative
Julie Sweetkind-Singer, NGDA Project
Myron Gutman, DataPASS Project
Victoria Reich, Stanford University

(B) *Rowing in the Same Direction – Collaboration across Disparate Organizations*
Panel Discussion
This is a Panel discussion that will first provide a short status report on the PeDALS project as well as touch on the responses we have received from government agencies and other external stakeholders.

As a group we will share the lessons that we have learned about effective collaboration internally, as well as across distance and political boundaries, including project agreements and efforts to sustain the project after the grant so that attendees can learn from our experience.

Lastly we plan to address the impact of creating a community of shared practice upon local practice and to explain how PeDALS fits into the partners' other digital projects and digital landscapes.

Presenters:
Matthew R. Guzzi, South Carolina Department of Archives and History
Abbie Norderhaug, Wisconsin Historical Society
Bonita L. Weddle, New York State Archives
Richard Pearce-Moses, Arizona State Library
Alan S. Nelson, State Library and Archives of Florida
Facilitator: Matthew Guzzi

3. <u>**Breakout Session Three: Collections and Storage**</u>
Ballroom

(A) *Challenges and Opportunities in Repository Interoperation*
Panel Discussion
The participants in this discussion panel will discuss recent work they have undertaken regarding challenges and opportunities in repository interoperation, including efforts to bridge LOCKSS and SRB preservation networks, and hub/spoke models for information exchange.

Presenters:
Martin Halbert, MetaArchive
Don Sutton, Chronopolis
Tom Habing, EchoDep Hub and Spoke project
Jonathan Crabtree, DataPASS Project
Facilitator: Gail McMillan, Virginia Tech

(B) *Promises of Preservation: What Do You Expect for Your Collections?*
Discussion Session
This session will focus on user requirements for storage of digital content collections, and why these requirements are important to solicit and quantify. The session is intended to be an interactive discussion about storage expectations and operational realities. Specific questions include:
- The challenges of identifying "best practices" in digital content storage
- How to apply "best practices" to NDIIPP network collections and similar types of collections.
- Collecting and quantifying user requirements for digital content storage for current and future collections.

The session will identify specific areas of interest or concern to NDIIPP partners, including new options such as "cloud computing" and solicit interest in future working groups or related projects.

Presenters:
Jane Mandelbaum, Library of Congress
Michelle Gallinger, Library of Congress

4. **Breakout Session Four: Digital Preservation of Dynamic Content**
   Hunt Room

(A) *Building a Digital Preservation System for Digital Motion Picture Materials*
Discussion Session

The purpose of this session is to provide information about the architecture, planning and challenges of building an interim digital preservation solution for digital motion picture materials. This session will discuss the AMPAS Interim Digital Preservation System. We will focus on the following issues:
- The Collection: DCI StEM Digital Materials
- The Content Audit Process
- The Hardware Architecture
- The Software Being Utilized
- The Role of Metadata
- Problems Encountered
- Next Steps

Presenter:
Nancy Silver, Academy of Motion Picture Arts and Sciences

(B) *Software Preservation and the OAIS Reference Model*
Paper Session

The University of Illinois Urbana-Champaign, along with the Maryland Institute for Technology in the Humanities, Rochester Institute of Technology, Stanford University and the University of Maryland, are conducting an investigation into the preservation of computer games and interactive fiction. While many within the digital library community have advocated the use of the Open Archival Information System Reference Model in designing digital preservation systems, an examination of the issues involved in the preservation of software indicates that there are problematic assumptions within the Reference Model with respect to the relationship between representation information and an archive's designated community, and the purposes which representation information will serve.   These assumptions may place software which is the object of preservation at risk unless libraries and archives implementing OAIS-compliant repositories give special attention to OAIS context information and the interplay between representation information, context information, and policies for collection development and management.

Presenter:
Jerome McDonough, University of Illinois

5. <u>**Breakout Session Five: Negotiations and Incentives**</u>
   Salon

(A) *Ingestion and Indigestion: Negotiating Between States in the Digital Cloud*
Discussion Session
Begun in 2007 under the leadership of the Washington State Digital Archives, the Multi-State Preservation Consortium has brought together 10 states and dozens of different agencies to create a centralized digital archive based on the Washington State model. Though highly successful the process has not always been smooth, and the lessons we have learned can be applied to other multi-partner digitization efforts. This discussion board will review the history of the partnership, outline some lessons learned along the way, and include time for audience questions and participation.

Presenter:
Kerry Barbour, Washington Secretary of State

(B) *GeoMapp Business Plan: Creating a Document to Get Stakeholder Buy-in*
Panel Discussion
In the current era of budget crises and program cutbacks, business justifications have become a critical part of developing and expanding digital archives. Come join the GeoMAPP team to discuss geospatial archives business plan and business case toolbox development.

Presenters:
Matt Peters, Utah Automated Geographic Resources Center, GeoMAPP
Glen McAninch, Kentucky Department of Libraries and Archives, GeoMAPP
Alec Bethune, North Carolina Center for Geographic Information and Analysis, GeoMAPP

6. <u>Breakout Session Six: Making Metadata Do Something Useful</u>
   Balcony

(A)   Access Points, Algorithms, Facets and Federation - – Making Metadata Do Something Useful
Presentations
Learn how to liberate your metadata and make it do something for you. No metadata is perfect, but there are good opportunities to use what you have. Facets, federation, algorithms, access points, the semantic web, and remediation are all good places to start. We'll focus on leveraging metadata for better access and visibility into digital content and collections. The session will start with four presentations from seasoned metadata wranglers, and will follow with a discussion to identify useful directions for future collaborations.

<div align="right">

Facilitators:
Jane Mandelbaum, Library of Congress
Butch Lazorchak, Library of Congress

</div>

(A1) Metadata and Preserving Geospatial Content
The role of geospatial and preservation metadata in the archival of spatial content. A primer on FGDC metadata, metadata cross walks, and the KY, NC, and UT approach to utilizing geospatial metadata in preservation activities.

<div align="right">

Presenters:
Matt Peters, Utah Automated Geographic Resources Center, GeoMAPP
Glen McAninch, Kentucky Department of Libraries and Archives, GeoMAPP
Alec Bethune, North Carolina Center for Geographic Information and Analysis, GeoMAPP

</div>

(A2) Minnesota States Metadata
Discussion on the Minnesota States Project's metadata scheme for legislative document interchange, use, and preservation, placing it in the context of the XML wrapper they've developed. Will touch on crosswalks and how the scheme can be generalized to work with other states' documents.

<div align="right">

Presenter:
Shawn Rounds, Minnesota Historical Society

</div>

(A3) Library of Congress Metadata for Digital Content
Description of LC initiative to develop standards and guidelines for digital content metadata.

<div align="right">

Presenters:
Rebecca Guenther, Library of Congress
Ann Della Porta, Library of Congress

</div>

(A4) NDIIPP Content Access Project - How Metadata is Used
"Best practices" for metadata as foundation of content access presentation developed with NDIIPP partner content.

<div align="right">

Presenters:
Jane Mandelbaum, Library of Congress
Uche Ogbuji, Zepheira

</div>

7. **Breakout Session Seven: Building Community and Partnerships**
    Hunt Room

(A) *Extensible Institutional Models for Digital Preservation: How Should Emerging Communities and New Content Domains Build Partnerships?*
Panel Discussion

As more scholarly communities awaken to the need for curation and preservation of specialized digital resources, these newcomers to the world of digital preservation have the opportunity to benefit from institutional developments emerging from the first generation of investments in digital preservation infrastructure. This panel will bring together senior staff from recently established digital preservation networks (i.e., MetaArchive, Chronopolis, ADPN) to discuss how newcomers can work with these partnerships: When is a given scholarly community or content domain well-suited to join an existing institutional arrangement? When might more institutions be needed? How can new scholarly communities leverage existing institutional arrangements to avoid duplication of effort? The panel will include a representative from the Web History Center, a new group that is seeking to implement a digital preservation solution. The conversation is therefore expected to explore specific and general solutions to the questions raised above.

Presenters:
David Minor, Chronopolis, UCSD
William Pickett, Web History Center and Rose-Hulman Institute
Martin Halbert, MetaArchive
Beth Nicol, Alabama Digital Preservation Network
Facilitator: David Kirsch, University of Maryland

8. **Breakout Session Eight: Tools**
  Ballroom

(A) *"Give us the tools, and we will finish the job": A Micro-Services Approach to Digital Preservation*
Panel Discussion
The California Digital Library (CDL) is deploying a new micro-services approach to digital preservation in which the range of functions traditionally encapsulated into a single repository system is devolved into a set of independent but interoperable services. Since each of the micro-services is small and self-contained, they are easier to develop, maintain, enhance, and, when appropriate, discard and replace. The initial set of micro-services is intended to address four key preservation goals: providing safety through redundancy, preserving meaning through interpretive description, satisfying utility through user-facing services, and enhancing value through use. This session will have two sections: the first will introduce the CDL micro-services approach and the tools and specifications underlying its implementation (e.g. N2T/noid, CAN/Pairtree/Dflat, BagIt/GrabIt); the second will provide an update on the JHOVE2 project (underlying the characterization micro-service), which is nearing its midway point. Progress to date includes a new processing model supportive of characterization at the level of individual files, byte streams arbitrarily-nested within files, and complex objects manifest in multiple files; signature-based identification; rules-based assessment; a redesigned core API; and a working prototype of the JHOVE2 command line interface.

Presenters:
Stephen Abrams, California Digital Library
Perry Willet, California Digital Library

(B) *Transfer Tools: The Other Outcome of "The Year of Content"*
Presentation Session
During 2008 the Library received over 80 TB of content from a variety of preservation partners. To accommodate these physical and network transfers the Library has developed the BagIt packaging specification, the BIL Java Library, and Parallel Retriever, VerifyIt, and Validation scripts to support the bagging and transfer of content. In addition, newly developed are the LocDrop deposit application and the Bagger GUI desktop tool, which have recently become available for limited partner testing, and an Inventory System recently put into production at the Library. This session will present the BagIt standard and all the tools, provide examples of uses by various partners and the Library, and discuss the open source release process on Sourceforge.

Presenters:
Leslie Johnston, Library of Congress
Brian Vargas, Library of Congress
Justin Littman, Library of Congress

9. **Breakout Session Nine: Access and Web Archiving**
   Churchill Room

(A) *Search and Access Technologies for Large Scale Web Archives*
Panel Discussion
Web archives can offer in the future unprecedented opportunities for information discovery and analysis due to the richness of their contents extending over long periods of time. However the effective search and exploration of long term web archives within a temporal context is a challenging technical problem due to its scale, and the heterogeneity, complexity, interlinked, and dynamic behavior of web contents. We will present the underpinnings of a scalable architecture and related technologies developed by the ADAPT group at the University of Maryland, which enable:

- Effective full text – temporal search and access of web archives whose returned results are evaluated within a temporal context;
- Grouping and summarization of returned results according to the user's specification;
- Temporal search that is substantially more effective than the "search and then filter" strategy;
- The possibility of highly parallel search and access strategies carried out on separate clusters.

This work has been carried out in collaboration with the Internet Archive and the Library of Congress, and will be demonstrated on a significant web archive provided by them.

Presenters:
Joseph JaJa, University of Maryland
Sangchul Song, University of Maryland
Mike Smorul, University of Maryland

(B) *Building Sustainable Digital Collections: Web Archiving for Public Access*
Presentation Session
Internet Archive has been steadily building collections and compelling content with a variety of valued partners who act as the content providers. In this program we will discuss building the partnerships, creating access, the tools and technology, lessons learned, some of the success stories and some of the challenges. Featured collections are developed by Virginia Tech University, Arizona State Archives, Library and Public Records as well as the Electronic Literature Organization and Stanford University.

Presenter:
Kristine Hanna, Web Archiving Services, Internet Archive

(C) *It's "Real:" Public Access to Archives Built With the Web Archiving Service*
Demonstration
In January 2005, the California Digital Library and partners at NYU and the University of North Texas embarked on "The Web-at-Risk", an NDIIPP grant to enable librarians and archivists to capture and preserve web-based government information. This resulted in the Web Archiving Service, which provides a full suite of tools for capturing and curating web sites. Until now, the archives created by curators have been "dark archives", with no public access. This has been a challenge for many current and potential curators for whom the service is not "real" until researchers can see the archives. In June 2009, the lights will go on, and the general public will be able to see the fruits of our labor. This session will demonstrate the Web Archiving Service public access features, using the archives of California state and local government agencies created by the project's curators.

Presenter:
Tracy Seneca, California Digital Library

10. Breakout Session Ten: Content Creators and Metadata
Salon

(A) *Content Creation and the Capture of Metadata: A Roundtable Discussion*
Discussion Session
Informal, off the cuff comparisons of the state of the art by content creators (and knowledgeable colleagues of creators) in still photography, video recording, geospatial data compilation, sound recording, and motion picture production. Segments of producer-created metadata support archiving and preservation, often by the creators themselves. Topics: the readiness of needed specifications, levels of adoption by practitioners, and levels of support by important software applications.

Presenters:
David Riecks, Stock Artists Alliance
Steve Morris, North Carolina State University
Nan Rubin, Preserving Digital Public Television Project
John Spencer, BMS/Chace
Nancy Silver, Academy of Motion Picture Arts and Sciences
Carl Fleischhauer, Library of Congress
Johanna Bauman, ARTstor
Richard Anderson, ASMP