

<b>Breakout Session Group One, Wednesday July 21, 2010, 2:45 p.m. – 4:00 p.m.</b>					
(1)	(2)	(3)	(4)	(5)	(6)
<b>The Original NDIIPP Partners: Plans for Moving Forward</b>	<b>Trustworthy Repository Criteria, Virtual Organizations, and Infrastructure</b>	<b>Web Archiving Tools and Access</b>	<b>Preservation Tools and Services: JHOVE2 and PREMIS</b>	<b>Organizational Relationships and Shaping the Digital Resource: Experiences from Four Partners</b>	<b>DuraCloud Pilot Program: Experiences of the DuraCloud Pilot Program: use cases tested and lessons learned</b>
<i>F. Scott Fitzgerald Ballroom A</i>	<i>F. Scott Fitzgerald Ballroom C</i>	<i>F. Scott Fitzgerald Ballroom D</i>	<i>F. Scott Fitzgerald Ballroom E</i>	<i>F. Scott Fitzgerald Ballroom B</i>	<i>Ernest Hemingway Salon 1 &amp; 2</i>
<b>Breakout Session Group Two, Wednesday, July 21, 2008, 4:15 p.m. – 5:30 p.m.</b>					
(7)	(8)	(9)	(10)	(11)	
<b>Politics of Sustainability: Making Distributed Digital Preservation Attractive</b>	<b>Modeling Interactions Between Repositories for the NDSA: Towards a Distributed and Collaborative Framework for Preservation</b>	<b>Challenges in Web Archiving</b>	<b>Data Integrity Monitoring in the Cloud</b>	<b>Case Studies from the Preserving State Government Information Initiative</b>	<b>DPOE Meeting (invitation only)</b>
<i>F. Scott Fitzgerald Ballroom A</i>	<i>F. Scott Fitzgerald Ballroom C</i>	<i>F. Scott Fitzgerald Ballroom B</i>	<i>F. Scott Fitzgerald Ballroom D</i>	<i>F. Scott Fitzgerald Ballroom E</i>	<i>Louisa May Alcott Boardroom</i>

## **Breakout Session Group One, Wednesday July 21, 2010, 2:45 p.m. – 4:00 p.m.**

### **(1) The Original NDIIPP Partners: Plans for Moving Forward**

*F. Scott Fitzgerald Ballroom A*

Nearly six years ago, the Library of Congress awarded the first round of National Digital Information Infrastructure and Preservation Program (NDIIPP) funding to eight institutions and their partners to identify, collect and preserve digital materials. Through the course of this funding, the original eight NDIIPP partners have defined and solved a wide variety of technical digital preservation challenges. At the same time, partners have defined and structured partnerships to coordinate workflow and insure project continuity.

This panel will discuss individual partners' plans to sustain themselves as they move beyond core Library funding. Discussion topics will include: What kind of structure and organization has their partnership established in preparation for moving beyond core Library support? How will the structure be different from the past six years? Have current economic realities affected the structure or organization of their partnership? What do they see as the biggest challenges for their partnership in the next six years? What types of future collaboration would they like to see with the Library of Congress?

#### **Presenters:**

Patricia Cruse, California Digital Library, Web-at-Risk project  
 David Kirsch, University of Maryland, Birth of the Dot Com Era project  
 Marc Maynard, University of Connecticut, Data-PASS project  
 Steve Morris, North Carolina State University Libraries, NCGDAP project  
 Nan Rubin, Community Media Services, Preserving Public Television project  
 Beth Sandore, University of Illinois at Urbana-Champaign, ECHO DEPOSITORY project  
 Tyler Walters, Georgia Institute of Technology, MetaArchive Cooperative

**Facilitator:** Abbey Potter, Library of Congress

### **(2) Trustworthy Repository Criteria, Virtual Organizations, and Infrastructure**

*F. Scott Fitzgerald Ballroom C*

Preservation of digital materials is one of the grand challenges of this century. In the field of preservation three fundamental approaches are converging. The first approach focuses on trust and checklists: identifying the criteria that digital repositories should satisfy in order to establish trust; providing a framework for documenting and auditing these criteria; and identifying organizations and programs that can act as certification authorities. The second approach involves partnership: where collaborating organizations and virtual organizations adopt joint stewardship of content or infrastructure. The third approach involves the development of digital infrastructure that provides integrated support for collaborative preservation.

This panel will reflect on the relationships among "trustworthy repository" criteria, virtual organizations, and infrastructure. What types of trust relationships are institutions creating through collaborative organizations? How should current trusted repository criteria be applied and to organizations and virtual organizations? In what ways does current infrastructure supporting digital preservation and access facilitate the creation and operation of collaborative preservation partnerships?

#### **Presenters:**

Micah Altman, Harvard University  
 Jonathan Crabtree, University of North Carolina, Chapel Hill  
 Nancy McGovern, University of Michigan, ICPSR  
 David Minor, University of California San Diego, San Diego Supercomputer Center

Katherine Skinner, Educopia Institute  
Mackenzie Smith, MIT Libraries

**Facilitator:** Barrie Howard, Library of Congress

### **(3) Web Archiving Tools and Access**

*F. Scott Fitzgerald Ballroom D*

This session will feature three presentations on web archiving and access.

**Facilitator:** Gina Jones, Library of Congress

#### *Memento: Time Travel for the Web*

During the plenary presentation, NDIIPP meeting attendees heard about what Memento is. This presentation will focus more on how to implement Memento in an existing archive, how to develop tools and clients, and other technical details.

**Presenters:**

Michael Nelson, Old Dominion University

#### *Web Archiving and Access*

Numerous sites have engaged in crawling and storing web-based collections using the standard WARC file format. However, much work remains to be done in developing tools to explore, manage, and provide easy and fast access to the crawled data. Unlike most collections, data in these collections have a temporal component, meaning they will likely be different at different points in the collections timeline. This session will cover tools and techniques for managing, viewing, and searching temporal collections. The University of Maryland ADAPT group has developed tools to harvest information from a large collection of WARC files while generating an interface to explore and discover information of the crawled web contents.

**Presenters:**

Joseph JaJa, University of Maryland  
Mike Smorul, University of Maryland

#### *A Flickr of Hope: Harvesting and Archiving Social Networking Sites*

The new “transparency and open government” movement sweeping the nation has led many government entities to begin using social networking tools as part of their communication strategy. One such user is the current Governor of North Carolina, Bev Perdue, who has requested that this material be archived as part of the public record. As a result, what has formed is a collaboration between the Governor’s Office, the State Archives, the State Library, and the Internet Archive aimed at making sure that all of the North Carolina government-based social networking sites are harvested and preserved. This presentation will discuss the technical challenges when harvesting social networking sites, each of the institution's methodology, lessons learned, current project status, and next steps.

**Presenters:**

Kate Odell, Internet Archive  
Jennifer Ricker, State Library of North Carolina

### **(4) Preservation Tools and Services: JHOVE2 and PREMIS**

*F. Scott Fitzgerald Ballroom E*

This session will feature two presentations on preservation tools and services.

**Facilitator:** Jane Mandelbaum, Library of Congress

*Next-Generation Characterization: An Update on the JHOVE2 Project*

JHOVE2 is a Java framework and application for next-generation format-aware characterization of digital objects. Characterization is the process of deriving representation information about a formatted digital object that is indicative of its significant nature and useful for purposes of classification, analysis, and use. This presentation will provide an update on JHOVE2 project activities and solicit feedback from the partners on project deliverables and plans for post-project sustainability.

**Presenters:**

Stephen Abrams, California Digital Library  
Tom Cramer, Stanford University  
Shelia Morissey, Portico

*Tools for implementing PREMIS metadata*

The second presentation of the session will show the new tools developed for implementing the PREMIS Data Dictionary for Preservation Metadata. The PREMIS in METS Toolbox was developed under contract to the Library of Congress by the Florida Center for Library Automation as a set of open-source tools developed to support the implementation of PREMIS in the METS container format. The session will also discuss the use of PREMIS controlled vocabularies from id.loc.gov (LC Authorities and Vocabularies Service) as part of a PREMIS tool.

**Presenter:**

Rebecca Guenther, Library of Congress

**(5) Organizational Relationships and Shaping the Digital Resource: Experiences from Four Partners**

*F. Scott Fitzgerald Ballroom B*

Four partners discuss "cat herding." The participants (or proposed participants) in the activities that the partners oversee range from photographers of fine art, creators of music and the record labels who publish and distribute their work, educators with a need for multilingual television programming, to gamers and electronic game creators. How do our partners enlist these varied individuals and communities in their activities, and how does (or will) this enlistment provide direct or indirect support for content preservation?

**Presenters:**

Johanna Bauman, ARTstor  
Maureen Droney, The Recording Academy  
Deborah Rossum, SCOLA  
Jerome McDonough, University of Illinois at Urbana-Champaign  
John Spencer, BMS/Chace

**Facilitator:** Carl Fleischhauer, Library of Congress

**(6) DuraCloud Pilot Program: Experiences of the DuraCloud Pilot Program: use cases tested and lessons learned**

*Ernest Hemingway Salon 1 & 2*

The DuraCloud pilot program commenced in the Fall 2009. DuraCloud is an open source technology for utilizing cloud infrastructure for preservation support and access services. The participants in this panel have all participated in a pilot program to test and deploy DuraCloud for managing content in a cloud environment. Each pilot partner will describe their use case, what problems they are trying to solve utilizing cloud technology, experiences and lessons learned during the pilot period. This summer the DuraCloud pilot partner program was expanded to include 13 new pilot partners to further test the software, and integrate with currently deployed production systems for managing content. The partners and use cases will be described briefly and future work planned by each current pilot partner will be discussed.

**Presenters:**

Chris Freeland, Center for Biodiversity Informatics, Missouri Botanical Garden  
Tom Garnett, Biodiversity Heritage Library, Smithsonian  
Michele Kimpton, DuaraSpace  
Peter Pinch, WGBH Media Archives  
Barbara Taranto, NYPL

**Facilitator:** Leslie Johnston, Library of Congress

**Breakout Session Group Two, Wednesday July 21, 2010, 4:15 p.m. – 5:30 p.m.**

**(7) Politics of Sustainability: Making Distributed Digital Preservation Attractive**

*F. Scott Fitzgerald Ballroom A*

Many distributed digital preservation projects are created and maintained by large institutions or organizations. This fact is not surprising given that several of the preservation alternatives are quite expensive and data needs are huge, but this approach can ignore the reality that many interesting and valuable digital resources are owned and / or created by smaller, less-well-resourced organizations. Opportunities for preservation must be made attractive to these organizations in order for a distributed model to be successful in the long run. This discussion will focus on models and strategies for achieving that end by engaging representatives from several existing distributed digital preservation networks in a conversation about the approaches deployed and outcomes accrued thus far. The conversation will also include some thoughts on how aligning the preservation systems themselves might offer a joint range of services to various customers.

**Presenters:**

Thomas C. Wilson, University of Alabama, Alabama Digital Preservation Network  
Ardys Kozbial, University of California, San Diego Libraries, Chronopolis  
Aaron Trehub, Auburn University, MetaArchive Cooperative  
Andrew Waller (in abstentia), COPPUL

**Facilitator:** Barrie Howard, Library of Congress

**(8) Modeling Interactions Between Repositories for the NDSA: Towards a Distributed and Collaborative Framework for Preservation**

*F. Scott Fitzgerald Ballroom C*

There is now great need for members of the digital preservation community to begin developing a new level of shared models and vocabulary for understanding inter - organizational content stewardship activities. Many NDIIPP initiatives have pioneered technical mechanisms (e.g., content replication and secure transfers) and governance structures (e.g., establishing curatorial and legal roles and responsibilities) that explicitly support collaborative practices across multiple, geographically dispersed preservation sites. As we enter the next era of stewardship activities under the National Digital Stewardship Alliance, how can we begin to document and share the lessons we have learned to date in this collaborative context? How can we use the principles that have emerged from this work to produce a critical and abstract understanding of the layers and dimensions of interaction that may take place not just within, but across preservation repositories?

This session will present an overview of collaborative distributed digital preservation efforts, including the progress that individual projects have made toward establishing early models for multi - repository networks and the benefits that the community of practice could gain from the development of a critical understanding of interactions between repositories in collaborative and distributed frameworks. We will

solicit the broader community’s engagement in beginning to build a taxonomy and abstract model for such shared repository environments.

**Presenters:**

Katherine Skinner, Educopia Institute  
 Martin Halbert, University of North Texas  
 David Minor, University of California San Diego, San Diego Supercomputer Center  
 Tyler Walters, Georgia Institute of Technology  
**Facilitator:** Michelle Gallinger, Library of Congress

**(9) Challenges in Web Archiving**

*F. Scott Fitzgerald Ballroom B*

As members of the International Internet Preservation Consortium (IIPC), libraries, archives, universities and other cultural heritage institutions around the globe have joined forces to enable the collection, preservation and long-term access of a rich body of Internet content. While members have benefited greatly from international collaboration in the development of tools, techniques and standards, each of our institutions experiences challenges related to archiving web content, some unique to our institutions, others common among across the IIPC membership.

Six members of the IIPC will provide updates on their overall framework and scope of their web archiving projects, and discuss some of the challenges they face.

**Presenters:**

Gildas Illien, National Library of France  
 Kris Carpenter Negulescu, Internet Archive  
 Mark Phillips, University of North Texas  
 Tracy Seneca, California Digital Library  
 Andrea Goethals, Harvard University  
 Martha Anderson, Library of Congress  
 Abbie Grotke, Library of Congress  
**Facilitator:** Abbie Grotke, Library of Congress

**(10) Data Integrity Monitoring in the Cloud**

*F. Scott Fitzgerald Ballroom D*

The advent of cheap remote storage through cloud services such as Amazon S3 and Rackspace offers an exciting new paradigm for storing and managing large scale data. However, most of these cloud computing sites present an opaque view of how your data is handled as a part of much larger volumes of widely heterogeneous data collections, and offer no specific guarantees on how the integrity of your data is managed. In this session, we will explore possible approaches for the data owner to monitor and manage the integrity of their cloud managed data in a cost effective way. A specific approach was developed by the ADAPT group at the University of Maryland, which is based on an extension of the ACE (Auditing Control Environment) enabling a data owners to remotely monitor and audit their data.

**Presenters:**

Joseph Jaja, University of Maryland (moderator)  
 Mike Smorul, University of Maryland  
 Andrew Woods, DuraSpace  
**Facilitator:** Jimi Jones, Library of Congress

## **(11) Case Studies from the Preserving State Government Information Initiative**

*F. Scott Fitzgerald Ballroom E*

This session will feature four presentations from each of the State Initiatives projects.

**Facilitator:** Butch Lazorchak, Library of Congress

### *Preservation and Access--Can One Live without the other?*

GeoMAPP team members will present different approaches for transfer of and access to geospatial information.

**Presenters:**

Kelly Eubank, North Carolina State Archives  
Matt Peters, Utah Automated Geographic Reference Center

### *Digital Archives: We have apps for that*

Three new applications have been developed to verify integrity, facilitate transfers, and ensure preservation of electronic records at the Washington State Archives, Digital Archives. These new tools work together to improve the ingestion and preservation of electronic records.

**Presenters:**

Justin Jaffe, Washington State Archives  
Dan Waterbly, Washington State Archives

### *Preserving Digital Legislative Content*

The project is working on three fronts: 1) education and promotion, to identify policies that shape programs and to raise the general level of awareness about the preservation of legislative content; 2) testing technologies, using MN, KS and CA as the guinea pigs, three different environments, with different possibilities and capacities; and 3) evaluation, looking at all that we've done, discussing costs/benefits, developing some guidelines, making some recommendations and providing some tools for analysis. Our general goal is to position legislatures to improve what they do, supporting a variety of partnerships, so users can share content, legislatures can share infrastructure and archives and libraries can share expertise.

**Presenter:**

Robert Horton, Minnesota Historical Society

### *Preserving Email: The PeDALS Approach*

The standards used in exchanging email suggest ways to automate the curation of archival email messages. The message headers offer a rich source of metadata to support discovery, and encoding standards for Internet transmission could serve as a basis for a submission information package. Appearances can be deceiving, though, and a number of such assumptions prove to be thorny in implementation. This presentation will discuss curatorial and technical aspects of three different methodologies the Arizona State Library, Archives and Public Record used to extract messages from Microsoft PST files, standardize metadata, and create AIPs and DIPs, as well as some of the surprises discovered along the way.

**Presenter:**

Pete Watters, Arizona State Library, Archives and Public Record

## **DPOE Meeting (invitation only)**

*Louisa May Alcott Boardroom*