

# Digital Preservation 2014 Program Guide

#digpres14

@ndiipp / @ndsa2

## Tuesday, July 22

---

**8:30 Registration and Continental Breakfast**

**9:00 Welcoming Remarks**

*Robert Dizard, Jr., Deputy Librarian, Library of Congress*

**9:20 National Digital Stewardship Alliance**

*Micah Altman, Director of Research Libraries, MIT, and Chair, NDSA Coordinating Committee*

**9:30 Software, It's a Thing**

*Matt Kirschenbaum, Associate Director of the Maryland Institute for Technology in the Humanities*

**10:00 Preservation Aesthetics**

*Shannon Mattern, Associate Professor, The New School*

---

**10:30 Break: Poster Session Preview**

Poster Session Co-Chairs, *Nicole Scalessa of The Library Company of Philadelphia* and *Michael Nelson of Old Dominion University*, will provide a short-preview of session, which is being held during the Digital Preservation Network (DPN) sponsored Reception, 5-7 pm in the foyer.

---

**10:50 "Stewarding Space Data" Panel**

*Hampapuram Ramapriyan, NASA*

*Deirdre Byrne, NOAA (NODC)*

*Emily Frieda Shaw, University of Iowa Libraries*

*Karl Nilsen, University of Maryland Libraries*

*Robin Dasler, University of Maryland Libraries*

*Vivek Navale, NARA (moderator)*

Terrestrial data presents significant digital preservation challenges, but the challenges of data from space are truly astronomical. In this session, four presenters will discuss and explore these issues in a range of contexts. These include projects working with satellite data from the 1950s and a database of extragalactic distances occurring in research libraries and ongoing work on managing space data and data from satellites at National Oceanographic Data Center and NASA. Exploring issues in curating space data in the last sixty years provides a fruitful context to better understand the lifecycle of research data in this, and other fields.

---

**12:00 Lightning Talks**

**Beyond the Russian Doll Effect: Reflexivity and the Digital Repository Paradigm**

*James A. Bradley, Ball State University*

Ball State University's Digital Media Repository (DMR) currently houses over 700,000 items in nearly 200 online collections (<http://libx.bsu.edu>). The DMR, traditionally considered a space for asset discovery and delivery, is beginning an exciting new phase which allows its assets to motivate new directions and spur creativity amongst users. This stimulus expands upon the original artifact to create new tangentially-related objects, which are then also added

to the DMR and linked in parity with the original asset that inspired it, creating a circular, reflexive, and spiraling model of growth. In this presentation, James Bradley provides tangible and practical examples of this paradigm by highlighting some current and future projects at Ball State University involving University Libraries' collaboration with the College of Architecture and Planning, Department of History, School of Art, and the David Owsley Museum of Art.

### **Video Game Source Disc Preservation**

*David Gibson, Library of Congress*

As the Library of Congress's custodial unit for video games, the Moving Image section receives roughly 400 games per year through the Copyright deposit process. A vast majority of these games come in as published works, complete with packaging and accompanying booklets. In some cases, however, the section receives master or source disks for the games, ostensibly an image of the game's complete file structure burned to a CD-R or DVD-R. It is often the case that these games were either unpublished in the United States in a physical form or never released at all. Using the never released PlayStation Portable game, *Duke Nukem: Critical Mass*, as an example, David Gibson will examine the various file formats and file structure of the source disc with an emphasis on preserving and providing access to the files and, ideally, a reconstruction of the complete game. Recommended for those who have an interest in software preservation and its inherent challenges.

### **UELMA-Compliant Preservation: Questions and Answers?**

*Rebecca Katz, Council of the District of Columbia*

The District of Columbia is considering how to implement a preservation model for an electronic publication of its code of laws. States implementing the Uniform Electronic Legal Material Act (UELMA) face a critical question: how to preserve electronic records to meet the preservation requirement of the law. The District of Columbia is looking at technological considerations, governance considerations, and basic content considerations. This presentation will highlight the decision points in the process of developing a preservation model and the factors playing into each decision made. How will we ensure that our data remains accurate? How will we maintain its accessibility to the public? How much will we preserve? Will we preserve only the text of the law, or the formatting as well? And how will we ensure the continued viability of a new system when its developers and champions are no longer in charge of the system?

### **We Want You Just the Way You Are: The What, Why and When of Fixity in Digital Preservation**

*Kate Murray, Library of Congress*

Fixity, the property of a digital file or object being fixed or unchanged, is a cornerstone of digital preservation. Fixity information, from simple file counts or file size values to more precise checksums and cryptographic hashes, offers data used to verify whether an object has been altered. Many in the preservation community know they should be establishing and checking the fixity of their content, but how, when and how often? The National Digital Stewardship Alliance Standards and Practices and Infrastructure working groups have published *Checking Your Digital Content: What is Fixity and When Should I Be Checking It?* to help stewards of digital objects answer these questions in a way that makes sense for their organization based on their needs and resources.

### **"Save Your Databases Using SIARD!"**

*Krystyna Ohnesorge, Swiss Federal Archives SFA*

For the archiving of relational databases, the Swiss Federal Archives have developed a format known as SIARD – Software Independent Archiving of Relational Databases. SIARD is an open standard and is supported by the SIARD Suite 1.0 application, which can be used to convert relational databases into SIARD format and can be ordered free of charge in conformity with the license agreement.

---

---

## 12:30 Lunch and Presentations: 2014 NDSA Innovation Awards Presented

### 1:00 Awards Presented

**Institution Award:** To acknowledge organizations taking an innovative approach to providing support and guidance to the digital preservation community.

**National Software Reference Library**, *National Institute for Standards and Technology*

**Doug White** accepting on behalf of the institution

Recognized for their substantial leadership in building a national collection of software, developing and sharing workflows and approaches for software preservation, and for modeling approaches to corpus analysis of born digital collections.

**Project Award:** To recognize projects whose goals or outcomes represent an inventive, meaningful addition to the understanding or processes required for successful, sustainable digital preservation stewardship.

**XFR STN**, *The New Museum*

**Johanna Burton** accepting on behalf of the institution

“XFR STN” (Transfer Station) acts as “an open-door artist-centered media archiving project.” The project is recognized in particular for how it stands as a model for community-driven digital preservation activity, utilized existing platforms like the Internet Archive for long term preservation, and worked directly with artist and media creators to preserve their work.

**Individual Award:** Given to individuals making a significant, innovative contribution to the digital preservation community.

**Trevor Muñoz**, *Assistant Dean for Digital Humanities Research at the University of Maryland Libraries and an Associate Director of the Maryland Institute for Technology in the Humanities*

Recognized for his work developing and teaching best practices in data curation in the digital humanities and for his work advocating for digital preservation as a core function of librarianship, archival work, and scholarship.

**Future Steward Award:** To acknowledge future stewards, especially students, but including educators, trainers, or curricular endeavors taking a creative approach to advancing knowledge of digital preservation issues and practices.

**Emily Reynolds**, *Program Specialist, Institute of Museum and Library Services*

Recognized for the quality of her work in a range of internships and student positions with ICPSR, University of Michigan Libraries, the Library of Congress, Brooklyn Historical Society, Storycorps, and, in particular, her recent work on the World Bank’s eArchives project.

---

---

## Afternoon Talk and Panel Discussion

### 1:40 Community Driven Innovation

*Cole Crawford, Executive Director, Open Compute Foundation*

### 2:10 “Community Approaches to Digital Stewardship” Panel

*Fran Berman, Rensselaer Polytechnic Institute*

*Bradley Daigle, University of Virginia*

*Aaron Rubinstein, University of Massachusetts Amherst*

*Jaime Schumacher, Northern Illinois University*

*Meg Phillips, NARA (moderator)*

The 2014 and 2015 National Agendas for Digital Stewardship both highlight the importance of coordinated community action in digital stewardship. This panel will feature brief presentations about cooperative approaches to digital stewardship from the Academic Preservation Trust, the Research Data Alliance, the Preserving Digital Objects

with Restricted Resources (PoWRR) and the Five Colleges Consortium followed by a panel discussion comparing and contrasting models for collaboration. The discussion will be moderated by Meg Phillips of NARA.

---

---

### **3:15 Break: Transition to Workshops and Breakout Sessions**

---

---

### **3:45 Workshops and Breakout Sessions I**

---

---

#### **Session 1: West End Ballroom – Salon C**

##### **The Fragile Surface: Preserving the CD-R**

*John Passmore, NY Public Radio*

Over the years more than 30,000 CD-Rs have been “burned” by WNYC producers, reporters and others to safeguard audio. Most of those discs (many with irreplaceable content) have ended in the vaults of the New York Public Radio Archives, where they have begun to show alarming symptoms, from occasional glitches to catastrophic failure. Consequently, one of the most recent formats in the Archives presents some of the most pressing preservation issues. The NYPR Archives embarked this year on a project to mass-transfer the content of its CD-Rs into the organization’s asset management system. To do so, the archives uses a suite of commercial tools and custom software to create automated workflows that both preserve the integrity of our data and provide accessibility both in-house and on-line to our listening audiences. This presentation will discuss the workflows that enabled the creation of thousands of digital preservation masters, while also capturing preservation data relating to materials, treatment, brand, batches, and number/severity of errors encountered during the ingest process.

##### **More than a Best Practice: Taxonomy as Tactical Preservation at NPR**

*Hannah Sommers, NPR*

NPR's mission is to create a more informed public, news cycle by news cycle, one story at a time. This forward-looking mission presents challenges for the preservation of our stories. NPR is not a cultural heritage institution but every day our team of Librarians steward cultural heritage. In 2014, taxonomy is a tool we use as a "stealth" preservation tactic targeting the digital storytelling that we consider to be at greatest risk. Ours is a story in progress. This presentation explores the challenges inherent in NPR's reality, and how our team has chosen to view our context as a set of creative constraints.

---

---

#### **Session 2: West End Ballroom – Salon A**

##### **Next Generation: The First Digital Repository for Museum Collections**

*Ben Fino-Radin, MoMa*

*Dan Gillean, Artefactual Systems*

*Kara Van Malssen, AVPreserve*

The Museum of Modern Art has for the past year worked with Artefactual Systems to design and build the first digital repository designed for the needs and uses of a museum collection of digital materials. The goal of this system is to assist time-based media conservators in providing sustainable collections care to variable, timebased media and digital artworks. Built upon the open source platforms Archivematica and Access to Memory (AtoM), the DRMC's foundation is in international digital preservation standards, digital collections care best practices, combined with cutting edge web application design standards and user expectations. This demo session will present one of the first public demonstrations of this new repository management software, and will center around discussion of its critical features, including: access to and comparison of quality file characterization, management and reporting of fixity checking, reporting and visualization of collections trends and anomalies, identification and collection of an artwork's technical dependencies, integration with TMS and MoMA's Digital Asset Management system.

## **Revisiting Digital Forensics Workflows in Collecting Institutions**

*Marty Gengenbach, Gates Archive*

This talk will provide updates on a series of 2012 interviews from archivists representing a variety of collecting institutions. Since conducting the initial interviews, there have been significant developments related to digital forensics and the acquisition of born-digital content: broader technological advancements including the release of the BitCurator tool set; in archival best practices, such as the adoption of EWF (Expert Witness Compression Format) over AFF (Advanced Forensics Format); and local institutional and organizational changes in technical infrastructure and personnel. The session will revisit questions from the original 2012 study: What digital forensics tools and practices have been implemented in archives? What challenges and successes have been encountered in the use of digital forensics applications? What trends can be identified across institutions using digital forensics applications? The session will also address challenges identified in the original research, and additional topics of interest to digital preservation practitioners.

---

---

## **Session 3: West End Ballroom – Salon E**

### **Research and Development for Digital Cultural Heritage: An Open Forum**

*Josh Sternfeld, NEH*

*Gail Truman, Truman Technologies*

As NDSA prepares to add new categories of content to its 2015 Agenda, including digital art and software, now is the ideal opportunity to assess the state of research and development for the preservation of digital cultural heritage. This break-out panel session will identify pervasive issues across formats found in digital cultural heritage that in time may lead to standards and practices adopted widely by practitioners working in museums, archives, libraries, arts organizations, and universities. Work in digital cultural heritage requires a creative balance of intellectual, theoretical, technical, social, and aesthetic matters. Building upon a spate of initiatives, conferences, and studies in recent years, this year's session will pose whether and how we can both embrace the novel properties intrinsic to each work or collection, while investigating the possibility of developing shared practices and standards.

---

---

## **Session 4: Dupont Ballroom – Salon G**

### **Beyond Transfer Stations: Collaborative Efforts Towards Digital Preservation of Analog Video and Born-Digital at XFR Collective**

*Andrea Callard, COLAB*

*Rebecca Fraimow, Dance Heritage Coalition*

*Julia Kim, New York University, Moving Image Archiving and Preservation*

XFR Collective is a New York City-based media preservation membership organization with a mission to preserve at-risk and obsolete audiovisual artwork by providing low-cost migration services to independent content producers for education, research, and cultural engagement. XFR Collective was formed to meet the needs of NYC-based arts communities. For a nominal membership fee, communities and artists can join the Collective to migrate content from increasingly obsolete media, both analog and born-digital, and make it publicly available on Internet Archive. The Internet Archive also serves as a repository for preservation-quality files. Panel themes include advantages of communal preservation initiatives, open access tools and resources, collaborative preservation strategies for research and scholarly use, and public access as a means of long-term preservation and sustainability.

Other collaborators on this project included: Ben Fino-Radin, Museum of Modern Art; Tara Hart, Jen Song and Johanna Burton, New Museum of Contemporary Art; Alexis Bhagat, Taylor Moore, Alan Moore, Coleen Fitzgibbon, COLAB; Leeroy Kun Young Kang; Walter Forsberg, National Museum of African American History and Culture; and Bleakley McDowell, New York University, Moving Image Archiving and Preservation.

## **Infrastructure Development: Multiple Digital Content Types in a Single Collection**

*Dina Sokolova and Jane Gorjevsky, Columbia University*

This presentation will focus on implementation of Columbia University Libraries' (CUL) digital repository infrastructure, and is based on an archival project involving a wide variety of digital materials. In 2011, CUL received a grant to permanently preserve and provide access to the archives of the Ford Foundation International Fellowships Program (IFP). The program offered fellowships for post-graduate study to social justice leaders from underserved communities in Asia, Africa, Latin America, Russia, and the Middle East who lacked access to higher education. One of the key objectives of the grant was to build a set of repository-based preservation, management, and access systems and services, which can later be utilized for other digital and hybrid collections. The presentation will focus on the project workflow and decision-making as well as technological tools, utilized in the process, within the framework of current digital preservation standards and best practices.

---

---

## **Session 5: Dupont Ballroom – Salon F**

### **The Great Migration: Moving First Generation Digital Texts to HathiTrust**

*Lance Stuchell and Kat Hagedorn, University Library, University of Michigan*

The University of Michigan has been scanning volumes locally since 1995 and making those volumes available through their original Digital Library eXtension Service (DLXS) since 2000. They recognized the usefulness of depositing their locally-scanned volumes into HathiTrust early on (in 2008), and deposited a large number of the volumes available only through DLXS into HathiTrust. In the last few years, they have been working methodically towards more stringent validation of ingested volumes for proper digital preservation, and more methodical ingest with an eye toward actual migration (rather than deposit and preservation in two places). Lance Stuchell and Kat Hagedorn will share their experiences with moving older digital files to an environment with more stringent technical preservation requirements, and how this experience can inform current views on migration and the preservation of legacy content.

### **Access and Preservation in the Cloud: Lessons from Operating Preservica**

*Jonathan Tillbury, Preservica*

The digital preservation community has recently been offered a series of cloud solutions providing various forms of functionality. In this talk Jonathan Tillbury describes a system that provides not just bit-level preservation but the full gamut of digital preservation services; up until recently, such systems were available only to organisations using a system installed on-site, created through a complex software development project. This “new paradigm” thus offers a zero capital cost “pay as you go” model to perform not just bit-level preservation but also “active preservation”. Importantly, it is also possible to export an organisation’s entire content thereby providing a suitable “end of system life” route.

---

---

## **5:00 Poster Session Presentations and Reception sponsored by The Digital Preservation Network (DPN)**

### **Digitization at USGS Libraries**

*Jenna Nolt, Qi Tong, U.S. Geological Libraries Program*

USGS Libraries have many unique historical materials in their collection that are essential scientific records for the earth science community. Some of these materials are in fragile condition and most of them are not available online. To preserve these materials and make them accessible online, the USGS Libraries embarked on a digitization effort. This poster will share the experience and challenges we have encountered to set up an in-house digitization operation. The poster will detail the different analog material formats; digitization equipment; process to upload digitized files to servers; training of staff, etc.

### **The Occupy Web Archive: Is the movement still on the live web?**

*Maria LaCalle and Scott Reed, Internet Archive/Archive-It*

With the spontaneous and rapid development of the Occupy Movement came a concerted effort to archive at risk web based content. In an effort to capture and preserve the thousands of new websites and blogs designed to announce and organize the movement, Archive-It created an "Occupy Movement" collection in November of 2011 and crawled over 900 URLs. This poster session will describe the URL selection, appraisal and collaboration process, providing access to the archived sites and offer an analysis of how many URLs archived during the project are no longer available on the live web.

[www.archive-it.org](http://www.archive-it.org)

### **Sustaining Digital Initiatives in the National Digital Stewardship Residency**

*Emily Reynolds and Julia Blasé, Former National Digital Stewardship Residents*

The poster discusses lessons learned in the National Digital Stewardship Residency program, where the authors engaged in project planning and implementation of digital archive initiatives. Several factors vital to sustainability for both projects will be described and explored, including: building organizational support, planning for future financial, technological, and personnel needs, and strategies for collaboration and resource-sharing in digital stewardship.

### **The Digital POWRR Project: Enabling Collaborative Approaches to Digital Preservation**

*Stacey Erdman, Northern Illinois University*

The Digital POWRR project has spent the last several years investigating how those from smaller or less resourced institutions can successfully engage in digital preservation activities within the constraints of their budgets and infrastructure. This poster presents an overview of the activities the team has undertaken, and presents abbreviated findings that will be part of the project's final white paper.

<http://digitalpowrr.niu.edu>

### **Picture This: Migrating Metadata in Visual Materials Collections**

*Savannah Gignac, Niels Bohr Library and Archives*

*Emilio Segrè, Visual Archives*

The poster documents the data migration of 25,000 image records from a historical photo database held in Microsoft Access, to a more flexible CMS platform. The presentation will highlight how we selected and designed a new CMS, evaluated existing metadata, prepped for data migration, and ensured better user access and search capabilities of our online photographic archive.

<https://photos.aip.org/>

### **Tyranny of Water**

*Ingi House, Defense Acquisition University and Progressive Technology Federal Systems*

This poster presentation will look at the challenges faced when trying to digitize a collection with various kinds of water damage. The first year of the recovery program is putting into place standards and protocols to insure that all information is captured, including evidence of the damage.

### **NOAA Digital Collections: Archiving, Online Access, and Metadata Exchange**

*Stanley Elswick, NOAA Central Library*

This poster gives an overview and highlights some of the unique digital collections, projects, and metadata exchange efforts the NOAA Central Library has created.

### **Web Archiving: Capturing Best Practices and Tools**

*Courtney Whitmore and Laura M. Gentry, Wayne State University Students*

Many information professionals are overwhelmed with web archiving and are seeking direction and advice on how to begin a web archiving project at their respective institutions. This poster presentation will give an overview of commonly used tools in web archiving and will offer miniature web archiving case studies as examples and lessons from the experienced professionals who are currently undertaking this task.

## **Enhancing and Promoting Data Services: CLIR/DLF Postdoctoral Fellowship Program in Data Curation**

*Rachel Frick, Council on Library and Information Resources / Digital Library Federation (CLIR/DLF)*

Launched in 2012 as an expansion of the CLIR Postdoctoral Fellowship Program in Academic Libraries, the CLIR/DLF Postdoctoral Fellowships in Data Curation have given recent PhDs opportunities to work in research libraries and data centers to help develop data management strategies, policies, tools, and services that relate to the fellows' own disciplines. This poster will describe the experiences and observations of the first two cohorts of these fellows, including CLIR/DLF Fellows in Data Curation for the Sciences and Social Sciences as well as the new CLIR/DLF Fellows in Data Curation for Medieval Studies.

<http://www.clir.org/fellowships/postdoc>

## **Digital Audiovisual Preservation and Data Tape Workflows**

*Lauren Sorensen and Paul Klamer, Library of Congress*

*John Spencer, BMS/Chace*

Media files are often large, complex and process heavy, resulting in the need for delivery on data tape from vendors, and many times storage on data tape. LTFS and TAR format create interesting questions for the archivist working to copy and maintain authentic data across systems. The poster examines two use cases, giving an overview of workflows, and the advantages and disadvantages from preservation perspective, specifically looking at audiovisual material and management of large files.

<http://www.americanarchive.org>

<http://www.bmschace.com>

## **Cariniana Network – a case study of a Brazilian Network of Digital Preservation**

*Laura Rezende, Federal University of Goiás and Brazilian Institute of Scientific and Technological Information (IBICT)*

*Sonia Cruz-Riascos, Fundação de Amparo à Pesquisa do Estado de Goiás (FAPEG)*

Assuming that the digital preservation needs to be a shared responsibility, especially the large volume of digital information being produced and the nature of digital technology, this poster aims to discuss the Brazilian Network of Digital Preservation - Cariniana. It's an initiative of the Brazilian Institute of Scientific and Technological Information (IBICT) to provide technology to promote the preservation of scientific, technological and cultural heritage, with the support of distributed computing resources, and from partnership with research institutions that are responsible for storing the data and physically integration by the use of LOCKSS (*Lots Of Copies Keep Stuff Safe*), an open source software developed by Stanford University with the purpose of preserving data in digital format.

## **Web Archiving in Practice**

*Jaime McCurry and Maureen McCormick Harlow, former National Digital Stewardship Residents*

This poster recaps some of the web archiving work that former NDSR residents Maureen McCormick Harlow and Jaime McCurry conducted for their NDSR residency projects. Maureen developed a thematic collection for the National Library of Medicine and Jaime engaged in outreach efforts related to the Folger Shakespeare Library's web collections. The poster addresses the entire web archiving process, from picking a theme to making the collections visible throughout an organization.

NLM Collection: <https://archive-it.org/collections/4254>

Folger Collections: <https://archive-it.org/organizations/576>

## **1.21 Gigahertz! Bringing Library Filmstrips Back to the Future**

*Lisa Peters and Zoe Stachel, University of Wisconsin-Madison School of Library and Information Studies*

The poster presents a collaboration between students from UW-Madison's School of Library and Information Studies (SLIS) and the SLIS library on a service-learning project for a Digital Curation course. The project involved digitization and digital preservation planning for the library's collection of rare 1960s educational filmstrips and accompanying audio recordings.

### **Digital Preservation Outreach: Student Projects at CUA**

*Lindsey Bright, Michelle Datiles and Sarah Ettinger, Catholic University of America*

Eleven students in the CUA Library and Information Science Digital Curation class (Spring 2014) worked in team and participated in four digital preservation outreach projects. The poster highlights the student project products.

### **Building an Archive from the Ground-UP: New Ways to Recruit Volunteers**

*Adam Mosseri, Wayne State University*

Trying to create a functional archive is no easy task. There are several factors that an archivist must consider when approaching such a challenge, like how will volunteers be recruited? How can I get graduate students and undergraduate students excited about archiving? After the volunteers have been recruited, how do you keep them interested to continue working on the project? Currently the archaeology archive at Wayne State University has accomplished this task. This poster will cover various aspects of how to approach an archive that is in need of attention, how to create interdepartmental cooperation in an academic environment and innovative ways to recruit undergraduate and graduate volunteers for archival and digital preservation projects.

### **Linked Data Registry: A New Approach to Technical Registries**

*Jonathan Tilbury, Preservica*

The challenge of maintaining a comprehensive file format registry, whilst central to effective Digital Preservation, has eluded the world community despite many initiatives. This poster shows a practical way to operate an international peer-to-peer network of contributors and consumers of technical registry information and is based on EU funded research under the ENSURE project.

[www.preservica.com](http://www.preservica.com)

### **Open Access to Government Data: Implications for Digital Curation**

*Cal Lee, University of North Carolina at Chapel Hill*

*Nancy McGovern, Massachusetts Institute of Technology*

*Alice Bishop, Council on Library and Information Resources*

*Suzie Allard, University of Tennessee at Knoxville*

In February 2013, the Office of Science and Technology Policy (OSTP) issued a memorandum to the heads of executive departments and agencies directing them to develop plans promoting open access (OA) to data and publications. This poster will summarize preliminary results and elicit feedback on a project to study the data curation needs implied by recent open access activities. We are conducting a structured content analysis of federal agency plans supporting open access to data and publications, interviewing leaders of projects recently funded by the IMLS to identify their experiences related to data curation needs, and placing these findings within the context of current data curation education offerings, caps and opportunities, as well as elaborating broader implications for cultural heritage institutions.

### **Efficient Thumbnail Generation for Web Archives**

*Mat Kelly, Old Dominion University*

A common method for presenting webpage change over time is to display a set of thumbnails of the mementos, or archived pages. In this work, we implement thumbnail summarization techniques that provide ease of browsing for captured mementos while optimizing the process needed to depict a representative summarization.

<http://ws-dl.cs.odu.edu/>

# Wednesday, July 23

---

8:30 Registration and Continental Breakfast

9:00 Workshops and Breakout Sessions II

---

---

## **Session 6: West End Ballroom – Salon C**

### **Digital Preservation Questions and Answers**

*Hosted by NDSA Innovation Working Group*

Come with your questions and we will do our best to answer them. For example, what should you do with an old SQL database? What skills should a job listing for a digital archivist include? When should you keep hidden system files around? What questions should you ask when acquiring digital photographs? These are exactly the kinds of practical and pragmatic questions that we face every day. In this interactive session, staff from NDIIPP and NDSA member organizations will field your questions in an open forum. Along with the face-to-face question and answer session, the session also formally launches Digital Preservation Questions and Answers (<http://qanda.digipres.org/>) an online platform for supporting exactly this kind of knowledge exchange on an ongoing basis. Participants are encouraged to ask and answer questions in this online forum alongside the session.

---

---

## **Session 7: West End Ballroom – Salon A**

### **Tools Showcase and Demo Session**

This breakout session will feature brief presentations and/or demos of tools, services or an application. Following the presentations, participants will have the opportunity to for informal, one-on-one or small group demo sessions with the presenters.

### **Wayne State University's digital collections infrastructure**

*Graham Hukill, Wayne State University*

At Wayne State University, we are in the process of building a digital collections infrastructure using Fedora Commons to store our digital objects, Solr as our primary search mechanism, and a combination of HTML / CSS / Javascript for the front-end user interface. These components are “glued” together with a Python based server that handles search, limited management of objects, and other specific functions. The missing piece in our digital collections infrastructure is a robust, reliable, sustainable, and easy-to-use tool for managing our digital collections and objects in Fedora Commons and Solr. This lightning talk will outline and demonstrate a python based tool we are creating for the management of our digital collections; a management and curatorial “glue”.

### **Community ScanDays: to create and maintain local personal histories**

*John Sarnowski, The ResCarta Foundation*

This demonstration will use open source software to digitally capture, annotate, index, store, discover, display and validate local individually owned materials. Individuals are welcomed to bring photographs, slides, letters, biographies, audio cassettes, and artifacts to their local library or historical society. Metadata is added to each digital file in Library of Congress standard METS/MODS/MIX/AudioMD. Individuals are given a CDR containing the digital files made from their analog materials. Individuals provide basic metadata on a form which includes a deed of gift which can be used to donate materials to the hosting institution if desired. Digital objects are gathered into collections, and indexed for inclusion into a community website, which is created the same day. The website is harvestable using OAI/PMH for inclusion into metadata stores like the DP.LA or display tools like ViewShare.org. The archive of objects can be validated over time by use of checksums. The contents or portions of the contents can be aggregated into state run digital archives. Software used in this demonstration will include Audacity, ResCarta Toolkit, jOAI and FileZilla.

## **Dance Heritage Coalition**

*Kat Bell, Dance Heritage Coalition*

Dance Heritage Coalition will feature a demonstration of the Dance Preservation and Digitization Project (DPDP). The DPDP offers preservation- and access-quality video digitization services, preservation storage, and a secure online access point to share the cultural record of dance. The community-oriented approach to digitization has helped save unique moving image records held in “Living Archives” of dance companies. With these regional digitization stations (or “hubs”) in New York City, San Francisco, and Washington, D.C., digitization goes hand-in-hand with outreach to the dance community to encourage records management and archival best practices. Preservation storage is offered to individuals, companies, and institutions that lack the infrastructure to be long-term stewards of digital files. The lightning talk will introduce the audience to the project, highlighting milestones from the past year. The demonstration will allow attendees to interact with the access point; learn about our digitization processes including technology and equipment specifications; see our metadata template and other resources used for education and outreach; and learn about fair use principles employed by the project.

<http://archive.danceheritage.org/>

## **AVPreserve Tools**

*Chris Lacinak, AVPreserve*

AVPreserve will be presenting the following three tools: COI, or The Cost of Inaction Calculator is a free web-based tool that helps organizations analyze the implications of waiting to digitize their audiovisual collections. COI is a counter-perspective to the concept of “ROI”, or Return on Investment and helps effectively articulate what is lost in the way of access, intellect and finances through inaction. Fixity is a free and open-source GUI-based application that helps organizations and individuals ensure the data integrity of their collections of digital files through automated review, documentation and reporting. MDQC is a free and open-source GUI-based application that reads the embedded metadata of a file or directory and compares it against a set of rules defined by the user, verifying that the technical and administrative specs of the files are correct. MDQC automates and minimizes the time needed to QC large batches of files, greatly increasing efficiency.

[www.avpreserve.com](http://www.avpreserve.com)

---

---

## **Session 8: West End Ballroom – Salon E**

### **Developing a Born-Digital Preservation Workflow**

*Bill Donovan and John Kearney, Boston College*

This presentation will focus on the research and development of best practices and protocols for dealing with born-digital archival materials, and putting these into action using a real-life test case -- an external hard drive given to Boston College's John J. Burns Library as part of the personal papers of the Irish soprano and harpist Mary O'Hara. They have employed a variety of software and hardware tools, including some traditionally used by law enforcement, to address the challenges of special collections libraries and archives as they seek to preserve electronic records for future generations. As these technical steps were being undertaken, multiple organizational, policy, and workflow questions emerged. They needed to differentiate what should be done immediately to protect and “clean up” electronic records versus what is more properly dealt with later by processing archivists.

### **BitCurator Project**

*Porter Olsen, MITH*

BitCurator is an effort to build, test, and analyze systems and software for incorporating digital forensics methods into the workflows of libraries, museums, and archives (LAMs). It is led by the School of Information and Library Sciences (SILS) at the University of North Carolina, Chapel Hill and the Maryland Institute for Technology in the Humanities (MITH) at the University of Maryland. While attendees may have heard about BitCurator before, this demonstration will showcase the significant advances made in the BitCurator environment over the last year. While alpha versions of BitCurator required the use of unwieldy command-line python scripts, the current iteration of BitCurator has an easy to use graphical user interface. Other features include the ability to mount disk images for data access, a duplication

removal tool, and PREMIS metadata capture and generation recording interactions throughout the disk image's lifecycle. The new interface and additional tools allow digital archivists with a range of technical expertise to incorporate digital forensics tools and methods in their digital preservation workflows. This lightning talk and demonstration will introduce these significant enhancements to the BitCurator environment for conference attendees to see and interact with first-hand.

### **Planning a Web Archiving Project**

*Kim Schroeder, Wayne State University*

The growth of web archiving tools is a blessing for those involved in saving today's cultural but temporal assets. Options range from simple open source tools to full service supported software. What are the capture results? All web archiving is not equal and after a careful 8 point quality assessment of the options, an evaluation has been developed to reflect the closeness of the "preserved" site to the actual current live experience. This assessment was then applied to more than 100 web sites for evaluation. Discussion will focus on defined quality, usability of the tools, quantification of the replication of the live experience, percent of original data preserved and stability issues. A chart of the results will be available for discussion. The intent is to serve as a tool for planning a web archiving project as well as managing expectations for the final result.

---

---

## **Session 9: Dupont Ballroom – Salon G**

### **Preserving and Rescuing Heritage Information on Analog Media**

*Fenella France, Library of Congress*

*Vivek Navale, NARA*

*Elizabeth Griffin, CODATA Task Group*

*Chris Muller, Muller Media Conversions*

The sciences and the humanities share a common need to access data from the past. Despite the amazing sophistication of modern (electronic) data and analyses, only older (analog) data can inform about past conditions reliably enough for accurate research into changes and evolution. Accessing heritage information correctly and comprehensively is a multi-faceted challenge, involving expert treatment of media such as photographic images, ageing papers and books, early digital devices (tapes, cylinders, optical disks), and confronting other problems with data that cannot be reproduced. By reviewing the complementary roles of archivist, librarian, educator, database manager and scientist, this session will emphasize the need for team work, and the inclusive approaches required so that our many heritage records of by-gone Nature and culture can be interpreted correctly for the benefit of humankind. Active participation is encouraged.

---

---

## **Session 10: Dupont Ballroom – Salon F**

### **A Model for Integrating the Publication and Preservation of Journal Articles**

*Kevin S. Hawkins, University of North Texas*

There are policy, technical, and workflow gaps in library efforts to preserve online journal literature. Since libraries are increasingly involved in journal publishing, HathiTrust, a shared preservation-quality digital repository, is a natural place to archive and provide access to journal literature to ensure its long-term preservation and discoverability. The University of Michigan Library is funding the creation of mPach, an open-source, end-to-end publishing system in which archiving in HathiTrust happens as a byproduct of publication rather than being carried out after the fact. The architecture of mPach, its envisioned workflow, and plans for creating a shared infrastructure for publishing open-access journals will be summarized.

## **Kairos: A Journal of Rhetoric, Technology, and Pedagogy**

*Cheryl Ball, Illinois State University*

Kairos: A Journal of Rhetoric, Technology, and Pedagogy is the premier online scholarly journal in digital writing studies, a field that researches technology's impact on writing and pedagogy. For two decades, Kairos has been an open-access, peer-reviewed journal read in 180 countries, and is the longest continuously running journal that publishes webtexts. The speaker and her team recently tried to move Kairos from a hand-crafted journal to a more sustainable editorial-management system, Open Journal Systems, and to populate the database, we mined metadata from the entirety of Kairos's back issues at the webtext level and media-element level (e.g., a webtext can average 25 files per piece, including HTML, JPG, PNG, MOV, CSS, etc.). They mined for a combination of 32 fields of metadata, using Dublin Core (as represented in OJS) and scholarly-multimedia-specific fields. Cheryl Ball will present the crosswalk between these three schema (DC, PKP, Kairos) and discuss several challenges to the project besides the obvious ones of time and labor.

---

---

### **10:15 Break**

---

---

### **10:45 Workshops and Breakout Sessions III**

---

---

#### **Session 11: West End Ballroom – Salon C**

##### **Preserving News and Journalism**

*Anne Wootton, PopUpArchive*

*Leslie Johnston, NARA*

*Aurelia Moser, Knight/Mozilla OpenNews*

*Edward McCain, Missouri School of Journalism*

As news is increasingly digital in nature, news organizations don't have the expertise or resources to properly archive or preserve their digital work for future generations — or even to preserve their work for a year from now. This is true at small regional newspapers struggling to maintain a web presence and keep digital backups of their articles, images, and videos. It's also true at the biggest national news organizations building interactive news apps and data-driven journalism projects. As journalists and archivists, we are alarmed. In this session, we will share stories from our varied experiences in digital news production and archiving. We come from different backgrounds, but we share a common desire to better unite journalists and news organizations with archivists and the archival communities best equipped to provide them with resources. This session is a continuation of initiatives that many of us are already involved in with colleagues from the Reynolds Journalism Institute, the Newseum, the Mozilla Foundation, the Internet Archive, the NYTimes, ProPublica, and the Washington Post.

---

---

#### **Session 12: West End Ballroom – Salon A**

##### **The National Digital Stewardship Residency Program in DC, NYC and Boston**

*Margo Padilla, Metropolitan New York Library Council (METRO)*

*George Coulbourne, Library of Congress*

*Andrea Goethals, Harvard Library*

*Vickie Allen, PBS*

*Nancy McGovern, MIT Libraries*

*Kris Nelson, Library of Congress*

*Jamie McCurry, Hillwood Estate, Museum & Gardens*

*Maureen McCormick Harlow, PBS*

The National Digital Stewardship Residency (NDSR) promises to bridge the gap between the theory and practice of digital stewardship by enabling qualified recent graduates the opportunity to participate in rigorous project work at established organizations. Through this collaborative opportunity, residents work to develop solutions to an array of

digital problems and challenges, enabling mutual learning and growth for both residents and the participating organizations. This discussion will center around topics related to being a host to a resident or replicating a NDSR program, including program administration, goals and structure, and ongoing program development.

---

---

### **Session 13: West End Ballroom – Salon E**

#### **Digital Preservation Audit and Planning with ISO 16363 and NDSA Levels of Preservation**

*Seth Anderson, AVPreserve*

*John Faundeen, USGS*

*Bert Lyons, AVPreserve*

The digital preservation community spent the previous two decades determining how best to create, preserve, and provide access to digital collections. These efforts led to numerous guidelines and standards outlining the functions and services of an optimal preservation infrastructure. For many organizations that have been developing preservation infrastructures according to the principles of these guidelines, the question has changed from *how* to *how well*. The introduction of criteria such as *ISO 16363 - Audit and Certification of Trustworthy Digital Repositories* and measurements such as the NDSA's *Levels of Digital Preservation*, provide structured approaches to determining the suitability and maturity of an organization's maintenance, security, and delivery of digital assets. This panel will demonstrate an assessment and gap analysis model that pairs the linear model of ISO 16363 with the horizontal metrics of the NDSA Levels of Digital Preservation. Members of the panel will present recent examples of audits and assessments utilizing these resources. They will illustrate the challenges raised by compliance with the standard and levels, and will examine whether certification is a necessity for establishing trust between an organization and its constituents.

---

---

### **Session 14: Dupont Ballroom – Salon G**

#### **Using Metadata to Support the Presumption of Authenticity**

*Adam Janesen, University of Hawaii*

As digital records are increasing being utilized in court systems, and therefore understood to a greater extent by judges and law makers, higher standards of conduct are being demanded of those collecting, processing and storing digital records used for evidentiary purposes. The result of these higher standards can be seen in the increasing amount of case law defining digital evidence collection and handling. As a result, there exists a stronger emphasis on developing mechanisms for digital records to able to demonstrate their authenticity, accuracy and accessibility over time. One such mechanism is imbedding an automated framework into the transfer process that records the ingestion of digital objects into a repository. This session will present one state archives approach to supporting a digital record's authenticity over time by integrating the METS and PREMIS metadata schemas into the transfer and processing of archival digital records in order to create a comprehensive set of documentation.

#### **Implementing Preservation Metadata: PREMIS Conformance and Best Practices**

*Evelyn McLellen, Artefactual Systems Inc.*

*Rebecca Guenther, Library of Congress*

Does your implementation of preservation metadata conform to the requirements of the *PREMIS Data Dictionary for Preservation Metadata*? If so, does it adequately capture all of the necessary functions of a digital repository that preservation metadata are supposed to support? Exactly what are those functions? This session presents the work of the PREMIS Conformance Working Group as it has tried to clarify and grapple with these questions over the past year. Evelyn McLellen and Rebecca Guenther present the work of the Conformance Working Group to date and invite the audience to voice their own questions and share their experiences using preservation metadata in their own institutional settings. Both experienced PREMIS users and newcomers to PREMIS are encouraged to attend.

---

---

## **Session 15: Dupont Ballroom – Salon F**

### **Project Pitching Session (sign-up required at the registration desk)**

*Funding Agencies: Institute of Museum and Library Services*

*National Historical Publications and Records Commission*

*National Endowment for the Humanities*

This session will offer the opportunity for attendees to meet one-on-one with the staff from three public funding agencies. Attendees may discuss ideas for project proposals or other topics of interest. **Sign-up for a 10 minute time slot discussion at the registration desk.**

---

---

## **12:00 Lunch and Presentations**

### **12:30 CURATEcamp: Digital Culture Preview**

*Amanda Brennan, Tumblr*

*Trevor Blank, The State University of New York at Potsdam*

*Trevor Owens, Library of Congress*

The web is a cultural platform. Across the web, a wide range of existing communities interact and express a diverse array of online communities have developed their own cultures. Unlike many other media, the participatory nature of the web has enabled a proliferation of the expression of these diverse cultures. As scholars increasingly turn to study this vernacular web, cultural heritage organizations responsible for collecting and preserving folklife and folklore need to develop plans and programs to collect and preserve records of these cultures and communities. Following Digital Preservation 2014, Trevor Owens of NDIIPP, Amanda Brennan of Tumblr and folklorist Trevor Blank will host a CURATEcamp unconference on this topic. In this session, they will each briefly describe the issues at hand and their goals going into the unconference.

### **12:40 Lightning Talks**

#### **Visual Haggard and Digitizing Illustration**

*Kate Halterhoff, Carnegie-Mellon University*

*Visual Haggard: The Illustration Archive* ([visualhaggard.org](http://visualhaggard.org)) is a digital archive intended to preserve, centralize, and improve access to the illustrations of popular Victorian novelist H. Rider Haggard (1856 - 1925). The majority of Haggard's approximately fifty novels including *King Solomon's Mines* (1885) and *She* (1887) were lushly illustrated, many of them repeatedly in different editions and by different illustrators. Illustration was always an essential part of reading Haggard's romances during the nineteenth-century. However, most recent published editions of Haggard's novels such as those from Penguin and Oxford Classics omit these visual texts. Illustration's absence and marginalization extends to digital archives. Google Books and Internet Archive's book scans often render illustrations haphazardly and in poor DPI quality. Absent illustration as well as grainy and low-resolution scans argue implicitly against the visual text's importance. This presentation will discuss the ways in which *Visual Haggard* uses digital preservation to re-prioritize the visual. The speaker will explain why focusing on visuality makes *Visual Haggard* a limited yet representative digital preservation project, and encourage others to consider illustration in their own work.

#### **DuraSpace and Chronopolis Partner to Build a Long-term Access and Preservation Platform**

*Michele Kimpton, DuraSpace*

DuraSpace and U of San Diego are building a robust, flexible Digital Preservation solution by integrating services DuraCloud and Chronopolis. The DuraCloud software and Chronopolis software are being integrated to provide a flexible end to end preservation solution for members of the Digital Preservation Network([www.dpn.org](http://www.dpn.org)). Some of the unique characteristics of this integration that could benefit Institutions archiving content in the Digital Preservation Network: any type of content can be ingested and stored; content can be transferred over the wire or via Internet 2; no special software needs to be installed by the Institution; content managed by Fedora and DSpace

repositories is seamlessly integrated with this DPN node; institutions can choose to keep their content in DuraCloud and a subset in DPN for the most flexible archiving solution; and no special packaging required by Institution, content is automatically packaged using “bag-it.” DuraSpace and UC San Diego will begin pilot trials this summer and start accepting production content the beginning of 2015. This lightning talk will highlight the key aspects of the project.

### **Library of Congress Recommended Format Specifications: Encouraging Preservation Without Discouraging Creation**

*Theron Westervelt, Library of Congress*

The Library of Congress has developed a set of hierarchies of physical and technical characteristics of creative works, which encourage preservation and long-term access. These will be used to help guide the Library’s own acquisitions, but the Library is also sharing them with external stakeholders for their benefit as well.

### **Electronic Records and Digital Archivists: A Landscape Review**

*Jane Zhang, Catholic University of America*

After the turn of the 21st century, more archival and other cultural heritage institutions start to hire/create positions for electronic records or digital archivists with responsibilities for collecting and processing electronic records and digital archival collections. This lightning talk presents and interprets the findings of a landscape review of the electronic and digital archivist workforce in the United States.

---

---

## **Panel Discussions and Afternoon Talks**

### **1:00 “Research Data and Curation” Panel**

*Elizabeth Yakel, University of Michigan, Ann Arbor*

*Ixchel Faniel, OCLC Research*

*Inna Kouper, Indiana University*

*Dharma Akmon, University of Michigan, Ann Arbor*

Effective data reuse depends on a complex chain of activities and involves the interplay of data producers, reusers, and curators. A key challenge is determining what specific data curation practices are needed to enable data reuse. For example, data undergo many transformations during their lifecycle and several people may contribute to these transformations. What should be captured, and how should it be captured so that other scientists can make use of the data? Even when data have been curated, reuse of data is not necessarily straightforward; there are important limitations to how easy data curation can make data reuse. In this session, four presenters will explore these issues in two different contexts: a cyberinfrastructure project that enables the transfer of active data into preserved data objects, and a case study of secondary uses of curated data in the archaeological community. A panel discussion will follow.

### **1:45 Contending with the Network**

*George Oates, Director, Good Form & Spectacle*

---

---

### **2:15 Break: Survey Announcements**

The *NDSA Standards and Practices Working Group* has created a short survey – only three required questions – to help us identify and rank some of the issues that may hinder digital video preservation.

The *2014 DPOE Training Needs Assessment Survey* is designed to capture perspectives on the continuing education needs of your organization in regards to digital preservation education needs.

---

---

**2:45 “Future of Web Archiving” Panel**

*Stephen Abrams, California Digital library*

*Martin Klein, Los Alamos National Laboratory*

*Jimmy Lin, University of Maryland*

*Michael Nelson, Old Dominion University*

The Web is celebrating its 25<sup>th</sup> anniversary in 2014, and although more organizations than ever are preserving web content, preservation of and access to archived content remains challenging on many fronts. This panel provides an opportunity to reflect not only where we are today, but the future of web archiving, featuring brief presentations by experts working in the field. Presenters will provide an update on the state of web archiving problems and opportunities, and discuss specific projects tackling some of the challenges. The panel and discussion following will be moderated by Stephen Abrams, California Digital Library.

**3:45 Big Data, Little Narration**

*Dragan Espenschied, Rhizome*

**4:15 Data Visualization and Analysis**

*Lev Manovich, Director, Software Studies Initiative*

**4:45 Wrap-up / Adjourn**

---

## Thursday, July 24

---

### **CURATEcamp: Digital Culture**

#### **The Catholic University of America, Columbus Law School**

The World Wide Web is a cultural platform. Across the web, a wide range of existing communities interact and express and a diverse array of online communities have developed their own cultures. Unlike many other media, the participatory nature of the web has enabled a proliferation of the expression of these diverse cultures. As scholars increasingly turn to study this vernacular web, cultural heritage organizations responsible for collecting and preserving folklife and folklore need to develop plans and programs to collect and preserve records of these cultures and communities.

This one day unconference will focus on exploring ideas and approaches to collecting and preserving digital culture both on and off the World Wide Web. It will be co-unchained by Trevor Owens, Digital Archivist at The Library of Congress, Amanda Brennan, Internet Librarian formerly of Know Your Meme currently at Tumblr, Trevor Blank, Folklorist who studies online culture.

This year’s camp is co-hosted by the Catholic University of America’s Department of Library and Information Sciences and will take place at the Catholic University of America, Columbus Law School in Washington, DC, July 24th, 9:00 am-4:00 pm.

See the CURATEcamp wiki for the agenda: [http://wiki.curatecamp.org/index.php/CURATEcamp\\_Digital\\_Culture](http://wiki.curatecamp.org/index.php/CURATEcamp_Digital_Culture)

The Signal blog post: <http://blogs.loc.gov/digitalpreservation/2014/06/register-for-curatecamp-digital-culture-july-24th/>