Digital Preservation Interoperability Framework (DPIF) Summary

Wo Chang; National Institute of Standards and Technology; Gaithersburg, Maryland, USA

Executive Summary

A two-part conference was conducted for the compilation and identification of requirements, technologies, and standards and best practices for the roadmap development of the long-term digital preservation interoperability framework. The US Workshop (co-hosted with NIST and INCITS at NIST, USA) was the first event focusing on US national needs and best practices. The second event, an international symposium (co-hosted with NIST and ISO/IEC SGDCMP, in Dresden, Germany) gathered different requirements and best practices at the international level. Submissions for both events were then combined as input (this document) towards the ISO/IEC SGDCMP for roadmap development and would be then used towards standardizing a digital preservation interoperability framework for effective and reliable access to preserved digital contents between interoperable digital preservation repositories.

1. Introduction

In October 2009, the International Organization for Standardization (ISO)/International Electrotechnical Commission (IEC) Joint Technical Committee 1 (JTC 1) reconstituted the Study Group (SG) on Digital Content Management and Protection (SGDCMP) with more focused terms of reference. Specifically, the attention of the SG was to target the issue of long term "digital preservation". This term was understood to include the protection and management of both digital data and digital content, including associated metadata. One of the SGDCMP 2010 tasks was to organize workshops and symposiums that would help develop a roadmap on long-term digital preservation standardization by identifying requirements, technologies, and best practices. Through the US International Committee for Information Technology Standards/Digital Content Management and Protection (INCITS/DCMP) and ISO/IEC SGDCMP, Wo Chang from the National Institute of Standards and Technology (NIST) was appointed as the program chair for both workshop and symposium events.

This was a two-part series on roadmap development for Digital Preservation. The US Workshop (co-hosted with NIST and INCITS, March 29-31 at NIST, USA), first focused on the US' national needs and best practices, and the second, was an international symposium (co-hosted with NIST and ISO/IEC SGDCMP, April 21 – 23 in Dresden, Germany), gathering different requirements and best practices at the international level. Both roadmaps would then be combined as input to the ISO/IEC SGDCMP to standardize a digital preservation interoperability framework for effective and reliable access to preserved digital contents between interoperable digital preservation repositories.

This paper presents some of a summary from the above two events. It is comprised of seven vision speeches from keynote speakers and 50 presentations categorized into the following three tracks:

- Content organizations (government, public/private institutes, etc.) for handling the preservation operations, strategies, and requirements,
- Technology developers (academia, commercial companies, R&D labs, etc.) for providing preservation approaches and solutions,
- Standards bodies (ISO/IEC, consortiums, industry associations, government initiatives, etc.) for establishing preservation best practices and standards.

2. Digital Preservation Interoperability Framework Events

The Digital Preservation Interoperability Framework (DPIF) events contain vital input for the roadmap development of the digital preservation interoperability framework. The input from these events will help us understand the current landscape of content organization operations, their requirements, the technologies used to solve these problems, and the standard and best practice procedures and approaches used to solve digital preservation problems in general.

US Workshop DPIF

The US Workshop DPIF had a main keynote speaker who addressed the overall goal of the US workshop on digital preservation interoperability framework. This was then followed by three visionary keynote speakers in three tracks: Content Organization, Technology, and Standards and Best Practices. Below are the 26 presentation titles in their respective tracks:

Content Organization Track

1. *The Interoperability Solution: Federated Search and Good Databases*, Dr. Walter Warnick, Director, U.S. Department of Energy, Office of Scientific and Technical Information
2. *NASA's Earth Science Data Systems - Lessons Learned and Future Directions*, Dr. Hampapuram Ramapriyan, Assistant Project Manager, ESDIS Project, NASA Goddard Space Flight Center
3. *NARA Electronic Records Archives Lessons Learned and Future Direction*, Mr. Dyung Le, Director, System Engineering, ERA Program, NARA
4. *LoC Repository Collections*, Mr. James Snyder, Senior Systems Administrator, National Audio Visual Conservation Center, LoC
5. *Archiving Strategy for USGS EROS Center and Our Future Direction*, Mr. John Faundeen, Archivist, Earth Resources Observation and Science Center, USGS
6. *Digital Goals and Challenges at the U.S. Government Printing Office*, Ms. Kate Zwaard, Lead Program Planner, FDsys Digital Preservation, U.S. Government Printing Office
7. *Safekeeping, Searching, and Reusing the Smithsonian, Institution's Digital Diamonds*, Ms. Isabel Meyer, DAMS Branch Manager, OCIO, Smithsonian Institution
8. *AWIPS Approach to Exponential Increase in NOAA Data Volume*, Ms. Deirdre Jones, Director, Systems Engineering Center, DOC/NOAA/NWS
9. *Improving And Strengthening Inter-Institutional Preservation*, Mr. David Minor, Head of Curation Services and Project Lead for Chronopolis, UC San Diego, San Diego Supercomputer Center
10. *The MetaArchive Cooperative: Chronicles in Cooperative Preservation*, Dr. Katherine Skinner, Executive Director, Educopia Institute

Technology Track

1. *Policy-based Data Management*, Dr. Reagan Moore, Director of the Data Intensive Cyber Environments Center, University of North Carolina, Chapel Hill, NC
2. *Analyses of Electronic Records: A Framework for Understanding File Format Conversions*, Dr. Peter Bajcsy, Research Scientist, NCSA/UIUC
3. *Balancing Performance and Preservation: Lessons learned with HDF5*, Dr. Mike Folk, Manager of the HDF Group, The HDF Group, Champaign IL
4. *Developing Bit Preservation Services at the Library of Congress*, Dr. Leslie Johnston, Digital Media Project Coordinator of Office of Strategic Initiatives, LoC
5. *Monitoring Distributed Collections Using the Audit Control Environment (ACE)*, Mr. Mike Smorul, Lead Programmer, UMIACS
6. *Lap Around the Windows Azure Platform - Scalable Compute and Storage Cloud Environment*, Mr. Vlad Vinogradsky, Sr. Principal Architect Evangelist, Developer and Platform Evangelism, Microsoft
7. *Open Data Protocol (OData) - Querying and Updating Data on the Web*, Mr. Vlad Vinogradsky, Sr. Principal Architect Evangelist, Developer and Platform Evangelism, Microsoft

Standards and Best Practices

1. *Curation Practices for the Digital Object Lifecycle, Part I: Determining Professional Competency Needs Through the DigCCurr Curriculum Development Project*, Dr. Helen Tibbo, Alumni Distinguished Professor, School of Information and Library Science, University of North Carolina, Chapel Hill, NC
2. *Metadata to support long-term preservation of digital assets: PREMIS and its use with METS*, Ms. Rebecca Guenther, Chair of PREMIS Editorial Committee, Library of Congress
3. *Long-term Digital Preservation for Medical Records*, Dr. Milton Corn, Deputy Director for Research and Education, National Library of Medicine, NIH
4. *Quest for an Advanced Image Standard: A Report From The Trenches*, Mr. Matthew Dougherty, Sr. Staff Researcher, National Center for Macromolecular Imaging/Baylor College of Medicine
5. *LOCKSS: Lots Of Copies Keep Stuff Safe*, Dr. David Rosenthal, Chief Scientist, LOCKSS Program, Stanford University Libraries
6. *The Washington State Model: Five Years of Digital Trials and Trails*, Mr. Jerry Handfield, State Archivist, Washington State Archives
7. *Towards Interoperable Preservation Repositories (TIPR)*, Mr. Joseph Pawletko, Software Systems Architect, Digital Library Technology Services, New York University
8. *DAITSS, an OAIS-based preservation repository*, Ms. Priscilla Caplan, Assistant Director for Digital Library Services, Florida Center for Library Automation

9. *Next Step: How Should Digital Preservation Interoperability Framework (DPIF) Go Forward?* Mr. Wo Chang, Manage of Digital Media Group, NIST

International DPIF Symposium

Despite numerous flight cancellations and airport closures from the Eyjafjallajokull volcano incident, the symposium managed to compress a three-day event into two days, utilizing webcast conferencing facilities for presentations and discussions. Below are the 26 presentations from this event:

Day 1 Presentation

1. *Developing a Digital Preservation Programme at a National Library*, Steve Knight, National Library of New Zealand
2. *NARA Electronic Records Archives Lessons Learned and Future Direction*, Wo Chang on behalf of Dyung Le, NARA, US
3. *ISO File Format for Digital Preservation*, Wo Chang, NIST, US
4. *The Usage of MPEG-21 Digital Items in Research and Practice*, Christian Timmerer, Klagenfurt University, Austria
5. *Digital Preservation: The Multimedia Standards way*, Mario Döller, University of Passau, Germany
6. *Introduction to MPEG-A Professional Archival Application Format (PA-AF)*, Noboru Harada, NIPPON TELEGRAPH AND TELEPHONE CORPORATION, Japan
7. *Stage 0 proposal of audio archive systems in IEC TC 100/TA7*, Noboru Harada on behalf of Kunimaro Tanaka, Teikyo Heisei University, Japan
8. *EuroVO Framework and Future AIDA Direction*, Francoise Genova, Strasbourg astronomical Data Centre, France
9. *Digital Preservation: Communicating Across Cyberspace and Time*, Kenneth Thibodeau, NARA, US
10. *PARSE Insight Framework and Lesson Learned*, David Giaretta, Science & Technology Facilities Council, UK
11. *CASPAR Framework and Lesson Learned*, David Giaretta, Science & Technology Facilities Council, UK
12. *Principles for Long-term Preservation of Digital Records*, Jim Suderman, co-investigator with the InterPARES Project, Canada

Day 2 Presentation

1. *Scientific Data e-Infrastructures in the European Capacities Programme*, Ms. Krystyna Marek, European Commission, Belgium
2. *Digital Archives for Molecular Microscopy*, Dr. Christoph Best, European Bioinformatics Institute-EMBL, UK
3. *The ESA Long Term Data Preservation Programme and Experiences Across CASPAR and GENESI-DR*, Dr. Vincenzo Beruti, European Space Agency, Italy
4. *The Planets IF - A Framework for Integrated Access to Preservation Tools*, Dr. Rainer Schmidt, AIT Austrian Inst of Technology, Austria
5. *The eXtensible Characterization Languages – XCL*, Dr. Manfred Thaller, Universität zu Köln, Germany
6. *LOCKSS & LuKII Project*, Dr. Michael Seadle, Berlin School of Library and Information Science, Germany
7. *The METAFOR project: preserving data through metadata standards for climate models and simulations*, Dr. Sarah Callaghan, British Atmospheric Data Centre, UK
8. *Policy-based Data Management*, Dr. Reagan Moore, Data Intensive Cyber Environments Center, University of North Carolina, Chapel Hill, NC, US
9. *Curation Practices for the Digital Object Lifecycle, Part II: Addressing Professional Competency Needs through the DigCCurr Professional Institutes*, Dr. Helen Tibbo, School of Information and Library Science, University of North Carolina, Chapel Hill, NC, US
10. *100 Million Hours of Audiovisual Content: Digital Preservation and Access in the PrestoPRIME Project*, Dr. Matthew Addis, Manager, IT Innovation, UK
11. *Quality Assurance: Towards Tools for Characterizing and Comparing Digital Documents*, Dr. Natasa Milic-Frayling, Microsoft Research Cambridge, UK
12. *Geo-Seas e-infrastructure*, Mr. Wo Chang on behalf of Colin Graham, Natural Environment Research Council (NERC), UK
13. *Digital Preservation: Lessons Learned Through National Action*, Mr. Wo Chang on behalf of Martha Anderson, LOC, US
14. *Singapore National Library Technical and Operation Challenges and Future Direction*, Mr. Wo Chang on behalf of Raju Buddharaju, National Library Board, Singapore

3. DPIF Events Summary – Keynote Speakers

Unsurprisingly, most of the content organizations presenting at both events already have mandates and are making their repository content available over the Internet. Due to the lack of standards however, specifically with various institutions' special needs and circumstances, they all agreed to use the Open Archival Information Systems (OAIS) reference model. The solution implementations for preserving content, however, vary. Below is a summary of each keynote speakers' talk for both events.

US DPIF: Main Keynote Speaker: Dr. Christopher Greer

Dr. Christopher Greer, Assistant Director for Information Technology Research and Development of the Office of Science and Technology Policy (OSTP) from the White House, discussed how the Obama administration's strategy for spurring American innovation and research development can be achieved through data access and preservation. Dr. Greer discussed two areas: (a) the establishment of an open government and (b) digital information preservation. Efforts to create an open government have allowed citizens to access federal agencies' activities via datasets and other important information online, fostering transparency and dialogue between the government and public. Yet creating this type of access is inherently linked with the need to further develop our efforts in the realm of data access. Regarding the preservation of digital information, four major challenges need to be addressed:

1. Information flow diversity: a 'one-size fits all model' does not work with communities of practices and a solution for developing higher levels of taxonomies that can integrate and aggregate information, including translation is crucial.
2. Data management expertise, time, and infrastructure: daily operation and management for preservation and access are critical. Within federal agencies alone, WorldWideScience.org, National Aeronautics and Space Administration (NASA)'s earth science project, Library of Congress (LOC)'s Audio and Visual Conservation Center, US Geological Survey (USGS) earth resources observation and science center, Government Printing Office (GPO)'s FDSys which makes government documents more easily accessible, Smithsonian Institution's digitizing its collections, NSF's and its DataNet activities, and the National Institute of Standards and Technology (NIST)'s Information Technology Laboratory on metrics and standards, all require the time, expertise, and resources to develop infrastructure that will ensure access and preservation.
3. Professional incentives: creating the right rewards and incentives framework is crucial. In 2005, a report from National Science Board on Long-Lived Digital Data Collections: Enabling Research and Education in the 21st Century lists the roles and responsibilities of individuals and institutions for sharing digital data collection among the data users and data authors.
4. Sustainable economic model: the ability to support the cost of long-term preservation. The Blue Ribbon Task Force led by Fran Berman and Brian Lavoie of Online Computer Library Center (OCLC) has done a remarkable job of bringing together data managers, data users, economists, librarians, and others to lay out principles for sustainable economic frameworks.

In order to accomplish the above, a policy should include a provision that projects and activities generating information must develop plans for managing data in the early stages of its development process. With that, Dr. Greer concluded his talk with two recommendations:

1. *Appropriate departments and agencies lay the foundation for agency digital scientific data policies and make them publicly available:* In laying appropriate policy foundations, agencies should consider all components of a comprehensive agency data policy, such as preservation and access guidelines; assignment of responsibilities; information about specialized data policies; provisions for cooperation, coordination and partnerships; and means for updates and revisions.
2. *Agencies promote a data management planning process for projects that generate preservation data:* The components of data management plans should identify the types of data and their expected impact; specify relevant standards; and outline provisions for protection, access, and continuing preservation.

US DPIF: Keynote on Content Organization Track: Dr. Kenneth Thibodeau

Dr. Kenneth Thibodeau, Director of Center for Advanced Systems and Technology at National Archives and Records Administration (NARA), presented the *Digital Preservation: Communicating across Cyberspace and Time.* In order to make a digital preservation interoperability framework work, it must focus on the following two paradigms:

1. *User Centric:*
    a. allow users to define collections with content from one or more repositories
    b. persist user-defined collections
    c. enable users to invoke services on collections from arbitrary providers
    d. support dynamic links between user-defined collections and repositories
    e. support appropriate use of data

2. *Generic Services Centric:*
    a. Repository/collection level services deal with identification, characterization, certification, assessment, interface specification, negotiation, and management tools
    b. Data Type Services deal with data type taxonomy, data format registration, data processing service taxonomy, format identification & verification tools, and format conversion tools
    c. Data Provenance Services deal with characterization of data provenance tools, provenance tracking across cyberspace, provenance enforcement, version control, and provenance tools
    d. Data Processing Services deal with Submission Information Package (SIP) and Dissemination Information Package (DIP) transmission, packaging tools, ingest tools, transformation tools, analytic tools, and preservation tools

Dr. Thibodeau concluded that the fundamental characteristics defining "infrastructure" for DPIF are: (a) independent of implementation, (b) evolvable, (c) extensible, (d) agile, and (e) interstitial.

US DPIF: Keynote on Technology Track: Dr. Sylvia Spengler

Dr. Sylvia Spengler, Program Director in Computer and Information Science and Engineering Directorate at National Science Foundation (NSF), gave a brief mention about NSF's mission established sixty years ago to support science and engineer and education that would promote the nation's well-being in both economic safety and health senses. NSF expects the investigators to share their data but the challenge is, under Bayh-Dole policy, the federal government may have a license-free use of the products of research but not other non-federal researchers. Furthermore, in many instances, replication of science requires more than the data, the operating system, the software application, and the network protocol the actual platform that things run on. If the operating environment is not available, it is hard to replicate the findings or prediction. For example, in the story of global climate change, you have different kind of observations: pictures of melting glaciers, numerical data on population fluctuations, satellite data for its surrounding, and training and testing of models. Data access and data interoperability are key to enabling the health of science.

NSF developed two programs: (a) Community of Practice program and (b) DataNet Program. The Community of Practice helps different scientific communities across the NSF programs at different stages of their development in understanding how they want to preserve and access their data at the next step. In some instances, a simple development of controlled vocabularies may be sufficient; others may be for ontology development, or true metadata acquisition through instruments and sensors. The DataNet program, on the other hand, enables scientific community to generate what we hope to see as new data organizations that will provide "user centric" data usage from varieties of sectors, federal, state, academic, and industry. The goal of DataNet is to provide a full lifecycle process from creation of data to dissemination of data by the original communities' consumption and by new communities for new discovery of data from other disciplines.

In conclusion, Dr. Spengler's hope is to find a creative technological approach in re-using scientific data within common theme of applications across different domains of applications.

US DPIF: Keynote on Standards & Best Practices Track: Dr. Francine Berman

Dr. Francine Berman, Vice President of Research at Rensselaer Polytechnic Institute (RPI), addressed *Building a Sustainable Foundation for the Information Age*. As co-chair for the Blue Ribbon Task Force on Sustainable Digital Preservation and Access, Dr. Berman laid out the tasks assigned for the task force: (a) conduct a comprehensive analysis of sustainable digital preservation (b), identify and evaluate best practices (c), make specific recommendations for action and, (d), articulate next steps for further work.

An important area discussed within the analysis was the economics of sustaining and preserving digital information and access; specifically, who would be paying for the preservation efforts. Government agencies, libraries, archivists, data users/owners/creators, or even private entities such as Google or Microsoft would not be able to single-handedly take on such

a task. Dr. Berman suggested drawing from today's economic models and defining key components for economic sustainability. Requirements include:

a) Recognition of benefits of long-term access and preservation
b) Incentives for decision-makers to act
c) Means of selecting "valued" information for long-term preservation
d) Mechanisms for supporting ongoing, efficient allocation of resources
e) Appropriate organization and governance of preservation and access activities

Alongside such recommendations, Dr. Berman also pointed out the current roadblocks:

a) Long-term preservation funded by short-term allocations
b) Lack of clear responsibility- "It's someone else's problem"
c) Misaligned incentives between those who can preserve and those who benefit
d) Challenges in prioritizing the value of digital preservation over more pressing short-term objectives

Dr. Berman further discussed the stakeholder's problems of *who benefits, who selects, who owns, who preserves, and who pay*s in four major digital content areas:

Data / Who	Research Data	Scholarly Discourse	Commercially Owned Cultural Content	Collectively-produced Web Content
Benefit	Greater research community	Greater research & learning community	General public, cultural historians	General public, cultural historians, etc.
Select	Often the individuals who generate the data	Publishers, based on community review	Studios, third-party organizations	Often the entities that preserve the data
Own	Often the data generators	Publishers generally own IP rights	Studios, third-party organizations	Often unclear
Preserve	Often the data generators and their proxies	Publishers and third-party entities	Institutional and individual repositories, third-party organization, etc.	Third parties interested in preservation of cultural assets
Pay	Federal agencies, institutions	Publishers, libraries, and third-party entities	Studios, professional organizations, private owners, custodial organizations, etc.	Third parties interested in the preservation of cultural assets

Dr. Berman concluded her talk with an action agenda for:

Funders and Sponsors of Data Creation

a) Create preservation mandates when possible
b) Invest in building/seeding stewardship capacities throughout the system
   - Fund the modeling and testing of domain-specific preservation strategies
c) Provide leadership in training and education for 21st century preservation, including domain expertise and core competencies in STEM.

Organizations

a) Fund internal preservation and access activities as core infrastructure.
b) Create economies of scope and economies of scale by partnering with related organizations and industry professional associations.

c) Develop preservation strategies that reflect technical, policy, and workforce best practices.

Individuals

a) Provide nonexclusive rights to preserve and distribute created content.
b) Partner with preservation experts throughout your data's lifecycle to ensure that data is ready to hand off in a form that will be useful over the long term.
c) Pro-actively participate in professional societies and relevant organizations to create stewardship best practices and selection priorities.

International DPIF: Keynote on Content Organization Track: Dr. Kenneth Thibodeau

Dr. Kenneth Thibodeau presented his visionary talk on Digital Preservation Communicating across CyberSpaceTime, similar to the US workshop (see above). He used a new metaphor of a motor-train transportation system between generations, expanding the cybergraphic information system model for re-use and re-purposing information object set through cyberspacetime.

International DPIF: Keynote on Content Organization Track: Ms. Krystyna Marek

Ms. Krystyna Marek, Program Officer for Digital Repositories (DR) at European Commission (EC), addressed the *Scientific Data e-Infrastructures in the European Capacities Programme*. The main thrust of the presentation was about the Scientific Data e-Infrastructure (SDI) domain launched in 2007 under the Capacities programme of Framework Seven (FP7). It complements the investment done by the European Commission, in a coordinated effort with European Union Member states to facilitate a rapid transition of e-Science among the European Members States that had made significant investments in e-Infrastructures for (a) linking ideas at the speed of the light, (b) sharing best computational resources, (c) accessing knowledge, (d) innovating the scientific process, and (e) experimenting in siloco. Three vectors of the renewed strategy were presented from the EC to the European Parliament and Council of Member States:

1. e-Science: defined as the combination of three different developments: the sharing of computational resources, distributed access to massive datasets and the use of digital platforms for collaboration and communication.
2. e-Infrastructure: integrates networks, grids and middleware, computational resources, experimental workbenches, data repositories, tools and instruments, and the operational support for global virtual research collaboration.
3. Innovation: by using fast networks and computational technologies, new discovery of knowledge may made possible from the existing information (usually, very large datasets of raw experimental data or curated data or even scholarly papers with research results).

Furthermore, there were the e-Science digital repository study recommendations:

1. Build an e-Infrastructure which ensures "research continuity" - funding, governance and management, leverage on other layers.
2. Engage users and service providers: support for data producers, trust and recognition, training and awareness.
3. Provide access to researchers, educators, and students - discovery and navigation, OA to publicly funded data
4. Maintain and preserve information - collections management, selection and appraisal for sustainability.

Before engaging in large scale e-Infrastructure initiatives, the EC decided to split the available resources for the SDI area in two open calls. The first call provided the opportunity for communities who already had well-identified challenges to propose projects addressing their most pressing problems related with data management. Selected projects covered the areas of bioinformatics protein databases, astronomic data, earth observation data, metadata for climate modeling and space-physics. The second call was designed to encourage communities to identify and tackle common boundary problems, which cut across different scientific domains, in order to promote multi-disciplinary collaborations integrating with other e-infrastructures domains of networking and computing. Below are the sample projects and their short descriptions in the SDI domain:

1. *NMDP:* Digital repository for cosmic ray data, pooling archives and collecting observations real-time
2. *IMPACT:* Improving protein annotation through coordination and integration of databases
3. *METAFOR:* Common Information Model and tools for using climate data and models
4. *GENESI-DR:* Open and seamless access to Earth science repositories (space, airborne, and in situ sensors data)
5. *DRIVER:* Flexible, robust, scalable, and cohesive pan-European infrastructure of Digital Repositories
6. *EURO-VA:* Virtual research environment for Environmental Monitoring and Fishery Resource Management

7. *PARSE.Insight:* Long-term preservation, permanent access to digital resources (intellectual capital of Europe)
8. *4D4Life:* Reengineering the 'Catalog of Life' (CoL), leading infrastructure in the field of taxonomy of living organisms
9. *D4Science:* Enables interoperability of data e-infrastructures in biodiversity, fisheries, and high energy physics
10. GEOSEAS: Access to marine geological and geophysical data from national geological surveys and research institutes
11. *SEALS:* Provides an infrastructure to allow the remote evaluation of semantic technologies
12. VAMDC – Works towards making the access to atomic and molecular data simpler and more integrated

International DPIF: Keynote on Standards & Best Practices Track: Ms. Martha Anderson

National Digital Information Infrastructure and Preservation Program (NDIIPP) was chartered in 2000 by the US Congress with the mission *"To ensure access over time to a rich body of digital content through the establishment of a national network of partners committed to selecting, collecting and preserving at-risk digital information"*. Over the past decade, NDIIPP has successfully formed a network of people and architecture for digital preservation. One can think of digital preservation as a race with relay teams proceeding in parallel lanes at different speeds towards the finish line. Enabling access across time and organizational and technical change is the equivalent of a finish line. Throughout the lifecycle of a digital object, there are numerous hand-offs across organizations and technical environments which need to pay attention to standards and best practices.

Below are the four major focus areas the NDIIPP worked on, including lessons learned:

1. Content: The NDIIPP partner projects have been working with images and text, audio visual, geospatial, and web sites. The work has been driven primarily by content type.

   Lesson Learned:

   a. Format standards are key to understanding how to manage digital content through changes in software and environments.
   b. Packaging standards are essential to moving content reliably between creator and stewardship entities. We learned this during our first project, Archive Ingest and Handling Test, and through transferring over 100 Tb of content across partner organizations.
   c. Metadata standards are important to content creators. We learned this from our partnerships with photography associations.

2. Organization: The NDIIPP partner organizations include international libraries, federal agencies, university libraries, state and local government, educators, cultural heritage organizations, media corporations, technical services developers, researchers, and specialized archives and collaborated with many initiatives.

   Lesson learned:

   a. Business processes are not interoperable. Sometimes it takes as much work to get the agreement and funding in place, as to do the project.
   b. Organizations need a local benefit to engage in larger collaborations.
   c. Organizations have interests beyond their own domain. Example: Photographers associations are interested in collaborating with museum imaging standards efforts.

3. Environment: The NDIIPP approach to infrastructure from a shared tools and services perspective, acknowledges that local resources would steer adoptions of storage and software at stewardship institutions.

   Lesson Learned:

   a. There is a high interest in sharing tools and services for preservation, but local repository adoptions depend upon organizational resources and policies.
   b. The simplest and most transparent tools are most often adopted. The broad acceptance of LOCKSS attests to this.
   c. Hardware vendors need to be engaged to understand preservation requirements. NDIIPP sponsors an annual storage meeting with vendors.

4. Communities: The NDIIPP efforts engage with many standards especially in the areas of images and text, audiovisual, geospatial, and web content.

Lesson Learned

a. Shared expertise can reduce time and effort to establish local preservation environments. Ex. The MetaArchive Cooperative
b. Working together builds trust.
c. Digital preservation is a task too challenging for a single institution. It requires multiple domain expertise about content and technology.
d. Communities are essential to sustain digital preservation efforts for the long-term.
e. Standards need to be considered in the context of diverse communities.

4. DPIF Events Summary – Track Speakers

The attached DPIF_Presentation_Summary.pdf lists summaries of the presentations/papers presented from both the US DPIF workshop and the international DPIF symposium. Summary questions were based on the Presenter Guidelines, which were divided into three tracks: Content Organization, Technology, and Standards and Best Practices:

Content Organization

- ***Background***
  a. High level description about institutions/collaborators providing given repository
  b. Describe what information the repository provides and if contribution is mandatory
  c. Describe digital content characteristics and who uses them

- ***Impact***
  a. Describe the given content's impact within the related field, or in other fields
  b. Describe what happens if the given content is lost
  c. Describe how digital preservation maximizes the value of the content

- ***Content***
  a. Describe content size (PB/TB/GB) and how often they get accessed
  b. Describe the scope of content collections and how they get distributed
  c. Describe what metadata, file format, packaging, and other related technologies are currently adopted, including rationale for such selection
  d. Describe system architecture for hosting the content
  e. Describe how 3d is related to OAIS reference model (if applicable)

- ***Access***
  a. Describe how users/systems can access your content, including rationale for any restrictions
  b. Describe what resources and capabilities (equipment, protocols, systems, expertise, etc.) are needed in order to access the content.

- ***Management***
  a. Describe what digital content needs to be preserved and for how long
  b. Describe preservation and lifecycle management strategies for the given content. This includes dealing with new technology adoption, outlining any changes in access anticipated during the preservation process
  c. Describe how strongly adopting other standards technologies as part of lifecycle management

- ***Protection***
  a. Describe content sensitivity including privacy, confidentiality, security, intellectual property and other rights

- ***Challenges***
  a. Describe barriers and lessons learned to make the repository available to the public
  b. Describe what preservation challenges your organization are facing
  c. Describe what tools and technologies you would like to see get develop

- ***Interoperability***
  a. Describe how repositories interoperability may benefit your organization

b. Describe how content sensitivity including privacy, confidentiality, security, intellectual property and other rights are affected under interoperability setting

Technology

- ***Background***
  a. Describe the motivation of the technology and what problems are being solved
  b. Describe the target audience using this technology

- ***Technology***
  a. Describe technology's novelty and innovative characteristics
  b. Describe technology's architecture and workflow
  c. Describe how such technology fits into existing digital preservation environment
  d. Describe technology deployment and feedback
  e. Describe technology specification documentation, support, and/or community activities

- ***Impact***
  a. Describe how the technology benefits to digital preservation community in terms of producer and consumer
  b. Describe productivity enhancement and operation improvement
  c. Describe what the potential cost saving with such a technology

- ***Development***
  a. Describe lessons learned when developing such technology
  b. Describe future development plan (if any)

Standards & Best Practices

- ***Background***
  a. Describe what standard body oversees the standard development and identify the participating members
  b. Describe the motivation of the standard and the problems being solved
  c. Describe who the target audience is that uses this standard or would be affected by the standard

- ***Standard***
  a. Describe standard's novelty and innovative characteristics
  b. Describe standard's architecture and workflow
  c. Describe how such standard fits into existing digital preservation environment
  d. Describe standard deployment, feedback and outreach into the relevant community

- ***Impact***
  a. Describe how such standards benefits the digital preservation community in terms of producer and consumer
  b. Describe productivity enhancement and operation improvement
  c. Describe the potential cost saving with such a standard

- ***Development***
  a. Describe lessons learned when developing such standards
  b. Describe future development plans (if any)

5. Acknowledgement

This paper presented a brief summary from the both DPIF events' keynote speakers and track speakers. Their knowledge, experience, insight are greatly appreciated in aiding the ISO/IEC SGDCMP's efforts for developing a roadmap on the digital preservation interoperability framework. The DPIF organizer would furthermore like to acknowledge and express appreciation towards organizations providing financial sponsorship such as Microsoft, the Library of Congress, the National Archives and Records Administration, and Intel for the US DPIF Workshop; and the Library of Congress and National Library Board of Singapore for their support in the International DPIF Symposium. These meetings would not have been made possible without their support.