How to Avoid Reinventing the Wheel in Preservation and Trustworthy Digital Repositories

Don Post
IMERGE Consulting, SNIA LTR
SavingTheDigitalWorld.org
PASIG Austin January 11, 2012

SAVING THE DIGITAL WORLD.ORG
An Alliance to Assure Future Access & Preservation of Digital Information
SavingTheDigitalWorld.org Goals

www.LTDPRM.org

- To be a strong advocate in North America for promoting long-term access and preservation of digital information
- To facilitate interchange of standards and best practices across participating organizations in different industries
- To encourage industry wide development of standards, best practices and knowledge necessary and sufficient to attain acceptable levels of persistent usability
- Scope: North America focus, but applicable worldwide
  - Europe is further ahead, e.g. funding and EU projects
  - US and Canada is a natural community of interest.
This is an outgrowth of past concerns about SNIA Long-Term Retention TWG and other initiatives.

SNIA is only one of dozens of organizations that are concerned with long-term digital information. Each has different perspectives and varied levels of credibility and influence on segments (subject matter, market, breadth).

Effective collaboration and consensus across participating organizations are needed to gain buy in and best results.

Has been incorporated as Not-for-Profit with Articles of Incorporation structured as a educational & charitable organization.
SavingTheDigitalWorld.org
Ties to Affiliated Organizations (through membership & interfaces)

- AIIM
- ARMA
- TAWPI-IFO
- SAA
- PASIG
- NAGARA
- And Several Others

- SNIA
  - LT Retention TWG
  - Cloud Archive & Preservation SIG
  - Data Protection & Optimization

- IEEE Computer Society
Conclusions About Urgency LTDP

● At the current rate, it is possible that in 20 years only 50% of critical info will be protected in TDRs
  – Much of the remainder will be lost forever
  – Many times worst than the 1973 disastrous fire at the NPRC destroyed 16-18 million files (OMPF)

● Now, it is virtually impossible to access knowledge from fragmented and specialized organizations dealing with LTDP. Today, learning from others is limited.

● There is a strong need for a comprehensive collaboration and links between fragmented organizations, knowledge and achievements
% of “Permanent” eRecords in 2010 that Are Still Accessible in Future Years:

Reasons:
- Inadequate EDRMS
- Policy & process issues
- Technology Issues

Digital information is in jeopardy: it is likely that in 35 years that 50% of critical info will be unavailable.
Avoid Reinventing the Wheel

- Standards based development and procurement is both necessary and partially sufficient to attain goals of Long-Term Access and Preservation.

- What are the standards and established best practices that you have experience with?

- “Home Grown” Trustworthy Digital Repositories may need to be replaced. Therefore TDR “A” must work interchangably with TDR “B” 20 years from now.
How to Avoid Reinventing the Wheel

Key Topics and Resources

- Learn from others across broad spectrum
- Best practices and technical standards
- Refer to LTDPRM.org bibliography SavingTheDigitalWorld.org/WIKI/Blog
  - International Standards
  - National Standards
  - Research and Demo Projects
  - Active Digital Archive Programs
  - Acknowledged Best Practices
- Why is it easier to reinvent the wheel rather than learning from others?
ITEM = Data, Document, or Record

Records “Declaration”
→ Retention, Disposition Rules

EDRMS PROCESSES
AIIM, ARMA
NAGARA

DIGITAL CURATION PROCESSES
SAA, ACA, NAGARA

STORAGE
SNIA, IEEE

Governments
Corporations

Libraries
Archives
Organizations with no EDRMS

Conventional Storage Processes
<10 yrs

TDR Functionality
to support 100 year access, preservation & usability

Long Term Retention TWG
Some Streams of Activity Related to Digital Curation

- Care & properties of physical media
- Digital forensics & data recovery
- Hardware & software interoperability
- Institutional & manuscripts archives
- Social science data archives
- Physical science data archives
- Librarianship (esp. digital libraries)
- Art & museum curation
- Medical information (e.g. imaging, informatics, health records)
- Lawyers & auditors
- Computer-supported cooperative work (CSCW)
- Management of information systems (MIS)
- Research on documents & document-centric computing
- Standards development
• American Association of Law Libraries (AALL)
• American Association for History & Computing (AAHC)
• American Association for State & Local History (AASLH)
• American Council of Learned Societies (ACLS)
• American Library Association (ALA)
• American Health Information Management Association (AHIMA)
• American Medical Informatics Association (AMIA)
• American Society for Information Science & Technology (ASIST)
• Association of American Publishers (AAP)
• Association of Canadian Archivists (ACA)
• Association of College & Research Libraries (ACRL)
• Association for Computers & the Humanities (ACH)
• Association for Computing Machinery (ACM)
• Association for Information & Image Management (AIIM)
• Association for Information Management Professionals (ARMA)
• Association for Library & Information Science Education (ALIS)
• Association of Moving Image Archivists (AMIA)
• Association of Public Data Users (APDU)
• Association of Research Libraries (ARL)
• The Association for Work Process Improvement (TAWPI)
• Audio Engineering Society (AES)
• Australian Society of Archivists (ASA)
• Business Forms Management Association (BFMA)
• CENDI Meeting (interagency working group of senior Scientific & Technical Information managers from nine U.S. federal agencies)
• Coalition for Networked Information (CNI)
• Collaborative Electronic Notebook Association (CENSA)
• Committee on Data for Science & Technology (CODATA)
• Corporation for National Research Initiatives (CNRI)
• Council on Library & Information Resources (CLIR)
• Computer Professionals for Social Responsibility (CPSR)
• Dance Heritage Coalition (DHC)
• Data Management Association (DAMA)
• Depository Library Council (DLC)
• Digital Curation Centre (DCC) Associates Network
• Digital Library Federation (DLF)
• Digital Preservation Coalition (DPC)
• DLM Forum
• European Commission on Preservation & Access (ECPA)
• Information Resources Management Association (IRMA)
• Institute of Certified Records Managers (ICRM)
• Institute of Electrical and Electronics Engineers (IEEE) – including IEEE Computer Society
• International Association for Social Science Information Service & Technology (IASSIST)
• International Committee on Technical Interchange for Space Mission Operations & Ground Data Systems (SpaceOps)
• International Council on Archives (ICA)
• International Federation for Information Processing (IFIP)
• International Internet Preservation Consortium (IIPC)
• Inter-Society Color Council (ISCC)
• Medical Library Association (MLA)
• Midwest Archives Conference (MAC)
• Museum Computer Network (MCN)
• National Association of State Chief Information Officers (NASCIO)
• National Association of Government Archivists & Records Administrators (NAGARA)
• Nuclear Information & Records Management Association (NIRMA)
• Online Computer Library Center (OCLC) – including Research Libraries Group (RLG)
• Open Grid Forum (OGF)
• Society of American Archivists (SAA)
• Society for Information Management (SIM)
• Special Libraries Association (SLA)
• Tape Head Interface Committee (THIC)
Common Difficulty in Sharing Industry Standards and Best Practices

Two organizations can easily share and discuss information
Common Difficulty in Sharing Industry Standards and Best Practices

Obtaining Information from multiple organizations is more difficult and time consuming.
Common Difficulty in Sharing Industry Standards and Best Practices

- Exchanging information between multiple organizations really starts to be complex.
- This is where conferences and associations help.
Common Difficulty in Sharing Industry Standards and Best Practices

• Bridging across industries and geographies is even more complex
Feasible Alternative Sharing Industry
Standards and Best Practices

Assembling the 2000 Piece Puzzle

• Building the Go-To Authoritative Knowledge Center

Well Designed Wiki with
Authoritative Subject Matter
Experts by Category
Feasible Alternative Sharing Industry Standards and Best Practices

- Bridging across industries and geographies is even more complex

<table>
<thead>
<tr>
<th>Topic/Categories</th>
<th>Topic/Categories</th>
</tr>
</thead>
<tbody>
<tr>
<td>INGESTION General</td>
<td>PRE-INGESTION General</td>
</tr>
<tr>
<td>INGESTION Policies</td>
<td>PRE-INGESTION Metadata Transformation</td>
</tr>
<tr>
<td>INGESTION File Formats</td>
<td>PRE-INGESTION Content Transformation</td>
</tr>
<tr>
<td>INGESTION Content Requirements</td>
<td>PRE-INGESTION Record Classification</td>
</tr>
<tr>
<td>INGESTION Metadata Reqmts</td>
<td>PERMISSIONS General</td>
</tr>
<tr>
<td>INGESTION SIP Definition</td>
<td>PERMISSIONS PE_Access Permissions</td>
</tr>
<tr>
<td>INGESTION Services</td>
<td>PERMISSIONS PE-Security &amp; Privacy</td>
</tr>
<tr>
<td>TDR SERVICES General</td>
<td>PERMISSIONS PE#LT Challenges</td>
</tr>
<tr>
<td>TDR SERVICES SIP-Ingestion Create AIP</td>
<td>EDRMS General</td>
</tr>
<tr>
<td>TDR SERVICES Object Transformations</td>
<td>DIGITAL General</td>
</tr>
<tr>
<td>TDR SERVICES Retention, Disposition, Hold</td>
<td>CURATION General</td>
</tr>
<tr>
<td>TDR SERVICES AIP Create DIP</td>
<td>TRUSTWORTHINESS General</td>
</tr>
<tr>
<td>TDR SERVICES AIP Preservation</td>
<td></td>
</tr>
</tbody>
</table>
AUTHORITATIVE KNOWLEDGE CENTERS

● AUTHORITATIVE
  – Power (n.) Influence Clout
  – Confidence (n.) Conviction Certainty Experience
  – Citation (n.) Source Evidence

● KNOWLEDGE
  – Information (n.) Information Facts
  – Comprehension Realization
  – Expertise Skill Know-How
  – Wisdom (n.) Education Intelligence

● CENTER
  – Consensus (n.)
  – Focus (n.) Concentrate
  – Place (v.) Arrange, Align
Welcome to the **Long-Term Digital Retention and Preservation Reference Model** site.

Help Define the Long-Term Digital Retention and Preservation Reference Model

While [OAIS](https://www.loc.gov/standards/oasis) exists as the ISO standard of a reference model for a digital archive, it doesn’t map adequately to the needs of large-scale distributed organizations operating on the internet or in the cloud and faced with the added cost burdens of legal and security risk, litigation, compliance and business requirements.

New capabilities and broader architectures are needed. Many organizations are extending OAIS, and it is time to begin gathering and defining best practices and common architectures. A simple example is all the noise around “Cloud Archiving.” We are leading the discussion, planning, using case definition, and best practices definition on the utilization of the Cloud for digital Archives. (Meaning preservation stores, not tiering.) 32% of the respondents in SNIA’s new survey declared that they plan on using the cloud for long-term retention periods more than 10 years. Sorry folks, traditional IT practices don’t support this. Significant portions of that data will be lost in 10 years. Is that acceptable?

[Join LTDP](https://www.loc.gov/standards/oasis) and help define these important areas.

The LTDP reference model community is being organized to respond to these many needs as digital preservation becomes an IT datacenter practice. Come join in and contribute. The time for development of cost effective, interoperable digital archives in the business and organization data centers has arrived.

[Join LTDP to Comment and Contribute - Register Here](https://www.loc.gov/standards/oasis)
LTDPRM.org Navigation

- News Feeds & Announcements >
- Discussion & Wiki >
- Solutions Guide >>
- Reference Model >>
- Reference Publications >>
- Resources >
- Calendar & Events >
- LTDP Glossary >
LTDP RM.org Navigation

- Solutions Guide
  - Buyer's Guide
  - Audit Methods
  - Consulting Resources
Reference Model
- Why We Need LTDP
- What is a Digital Archive?
- Preservation Services
- Preservation in the Cloud
- Preservation System Architecture
LTDDPRM.org Navigation
Long-Term Digital Preservation Model

- Reference Publications
  - Academic, Library, Archives Publications
  - Association Publications
  - Governmental, Standards, and Agency Publications
  - Vendor Publications
  - Conference Proceedings
  - Digital Preservation Surveys and Research
  - WebCasts-Videos
Registration & Sponsors

- About & Contact
- Join LTDP
  - Register Here
  - Sign in Here
- Sponsored By:
  - www.imergeconsult.com
  - http://www.ilm20.org/
What is Digital Preservation?

- Regardless of Source of Digital Preservation Objects –
  - Records Management
  - Digital Curation
  - Creation of Archival Object
    (e.g. SIP, AIP, TDO, SIRF Object)

- Digital preservation is a complex system of managed activities and a supporting infrastructure that can ensure digital records and information assets are accessible, useable, and trustworthy for as far into the future as is necessary.