Digital Preservation

UN FAO
May 23-24, 2011
Tom Cramer
<table>
<thead>
<tr>
<th>Time</th>
<th>Session</th>
</tr>
</thead>
<tbody>
<tr>
<td>08:30 – 08.35</td>
<td>Review of objectives and agenda</td>
</tr>
<tr>
<td>08:35 – 09:30</td>
<td>Infrastructure and tools</td>
</tr>
<tr>
<td>09:30 – 10:30</td>
<td>Case study: preservation activities at CDL</td>
</tr>
<tr>
<td>10:30 – 11:00</td>
<td>Morning break</td>
</tr>
<tr>
<td>11:00 – 12:00</td>
<td>Case study: preservation activities at Portico</td>
</tr>
<tr>
<td>12:00 – 12:30</td>
<td>Preservation initiatives and organizations: DataNet, DCC, DPC, IIPC, NDSA, OPF</td>
</tr>
<tr>
<td>12:30 – 14:00</td>
<td>Lunch</td>
</tr>
<tr>
<td>14:00 – 15:00</td>
<td>Case study: preservation activities at Stanford</td>
</tr>
<tr>
<td></td>
<td>Other preservation resources</td>
</tr>
<tr>
<td>15:00 – 15:30</td>
<td>Afternoon break</td>
</tr>
<tr>
<td>15:30 – 16:00</td>
<td>Format characterization</td>
</tr>
<tr>
<td>16:00 – 16:30</td>
<td>Characterization in preservation workflows</td>
</tr>
<tr>
<td>16:30 – 17:00</td>
<td>Questions and discussion</td>
</tr>
</tbody>
</table>
Preservation Initiatives & Organizations

DataNet
DCC
DPC
DPE
IIPC
NDSA
OPF
DataNet

• Sustainable Digital Data Preservation and Access Network
• Project from the US National Science Foundation – Office of Cyberinfrastructure
• $100,000,000 for five projects envisioned
• Goal = produce a network of cyberinfrastructure partnerships for scientific data deposit, analysis, preservation and use
Digital Curation Centre (DCC)

- Leading UK institution focused on research data
- Consortial effort, sparked by JISC in 2004
- Data Curation = Management across the information lifecycle (creation, selection, enrichment, preservation, access)
- http://www.dcc.ac.uk/
Digital Curation Centre (DCC)

- International Journal of Digital Curation (IJDC)
- Curation Lifecycle Model
- Training, White Papers
- Policy Analysis & Advocacy around Data
- Data Management Plans
Digital Preservation Coalition (DPC)

- Coordination, education and advocacy for digital preservation in the UK
- Founded in 2001, member-based organization (12 full and 24 associates)
- Introduction to and training on DP
- Technology Watch Reports
- http://www.dpconline.org
Digital Preservation Europe

• DP collaboration, awareness, service development and capacity building in Europe.

• Founded and funded as part of EU’s Sixth Framework Programme (FP6), 2002 – 06
  • Emphasis on coordinated research in EU
Digital Preservation Europe

- Reports
- Registries for trainers, training materials, projects, repositories, et al.
- DRAMBORA audit tool
- Platter planning tool
- http://www.digitalpreservationeurope.eu/
International Internet Preservation Coalition (IIPC)

- Preservation of Internet content
- Founded in 2003, with largely (but not only) national libraries
- 40 members worldwide
- Software & Tool development
- Working groups (e.g., Harvesting, Access)
- Conferences & workshops
- http://www.netpreserve.org
(US) National Digital Stewardship Alliance

• Coordinated by the Library of Congress
• Collaboration and exchange forum
• Founded in 2010, as extension of the NDIIPP program (2000 – 2009)
• 75 members – US Based
• Five working groups:
  – Content, Standards & Best Practices, Infrastructure, Innovation, Outreach & Education
Open Planets Foundation (OPF)

- “Practical solutions & expertise” in digital preservation, building on PLANETS project
- 16 members – Largely European
- http://www.openplanetsfoundation.org/
Open Planets Foundation (OPF)

- Plato – Preservation Planning Tool
- Testbed & Test Corpora
- Fido – Python implementation of DROID
- Format Registry Editing Tool
- OS Collaboratory – VCS, wiki, issue mgmt,
Other Preservation Resources

EPrints
DuraSpace
- Fedora
- Dspace
- DuraCloud
Licensing
LOCKSS
**EPrints**

- Turnkey open access and institutional repository software
- Run out of University of Southampton (UK)
- First deployed in 2000
- 269 known adopters worldwide, heavy UK user base
- Currently at Version 3.1.3
- Adding preservation services
DSpace

• “Digital Repository in a Box”
  – Quick to deploy
  – Less flexible data model and code base
• First deployed in 2002
• Estimated 1000 adopters worldwide
• Currently at Version 1.7
• Currently working on DSpace with “Fedora Inside”
Fedora

• Flexible Extensible Digital Object Repository Architecture
• Digital Repository Middleware – provides strong object modeling framework and hooks to services
• Research started in 1997; Version 2.1 released in 2005
• Estimated ~200 adopters worldwide
• Currently at Version 3.4
DuraSpace, DuraCloud

• Parent organization of DSpace and Fedora Commons
  – Merged in May 2009
• Supported by grants, sponsorship and service providers
• DuraCloud: Service broker for cloud-based storage and repository services for back up / DR, preservation, image & video access
Licensing

• Copyright = “All Rights Reserved”
• Creative Commons = “Some Rights Reserved”
• Creators can explicitly ok conditional use of content
  – Attribution
  – No Derivatives
  – Non-Commercial
  – Share-Alike
  – Public Domain
LOCKSS

- Lots of Copies Keeps Stuff Safe
- Originated at Stanford University
- Peer-to-peer, decentralized digital preservation system
- Focus is on scholarly articles
  - 7100 e-journal titles, 470 publishers
  - Collects web-based content
  - Preserves it locally
  - Provides 100% post-cancellation access
  - Done with publisher permission
LOCKSS

Capture & Replication
LOCKSS

Audit & Healing
LOCKSS

• Commodity Hardware & Open source software & Appliance = very low cost
• Follows traditional model of library-based distribution and preservation
  – Lots of Copies
  – Locally Managed Copies
• Publisher permissions ensure legal coverage
• Extensible to other collections
LOCKSS

• CLOCKSS: Controlled LOCKSS
  – Not-for-profit archive for ensuring access to orphaned scholarly content
  – One dozen major publishers + libraries

• Private LOCKSS Networks
  – Alabama Digital Preservation Network
  – Arizona State Library, Archive & Public Records
  – Council of Prairie & Pacific University Libraries Consortium
  – Data Preservation Alliance for the Social Sciences
  – Digital Commons – Berkely Electronic Press
  – MetaArchive Cooperative Project
  – Digital Federal Depository Library Program