

# Storage Architectures for Digital Preservation

## September 22-23 2009

### Washington DC

**Background Reading for Attendees**  
(These items are not intended as exhaustive or prescriptive; they are representative of material and projects that might be useful for context related to the meeting topics)

#### Library of Congress

- **National Digital Information Infrastructure Preservation Program (NDIIPP)**  
<http://www.digitalpreservation.gov/library/>  
“The mission of the National Digital Information Infrastructure and Preservation Program is to develop a national strategy to collect, preserve and make available significant digital content, especially information that is created in digital form only, for current and future generations.”

#### Digital Preservation Needs and Requirements

- **The Digital Dilemma**  
(Science and Technology Council, Academy of Motion Picture Arts and Sciences)  
<http://www.oscars.org/science-technology/council/projects/digitaldilemma/index.html>
- **Mind the Gap: Assessing Digital Preservation Needs in the UK**  
<http://www.dpconline.org/docs/reports/uknamindthegap.pdf>  
“Throughout this report, we have used the term ‘digital preservation’ to refer to all the actions required to ensure ongoing, meaningful access to digital information for as long as it is required and for whatever legitimate purpose.”
- **Sustaining the Digital Investment: Issues and challenges of economically sustainable digital preservation: Interim Report of the Blue Ribbon Task Force on Sustainable Digital Preservation and Access**  
<http://brtf.sdsc.edu/>

#### Practices and Practitioners

- **Digital Preservation: An Historical Overview**  
(Peter Verheyen, Syracuse University)  
(includes a reasonable bibliography of basic resources)  
<http://www.philobiblon.com/upload/IST676-DigitalPreservation.pdf>

- **How are We “Ensuring the Longevity of Digital Documents?”**  
(David Rosenthal)  
  
<http://blog.dshr.org/2009/04/spring-cni-plenary-remix.html>
- **Storage Networking Industry Association (SNIA)-Data Management Forum: Building a Terminology Bridge for Digital Information Retention and Preservation Practices**  
[www.snia.org/.../SNIA-DMF\\_Building-a-Terminology-Bridge\\_20090515.pdf](http://www.snia.org/.../SNIA-DMF_Building-a-Terminology-Bridge_20090515.pdf)
- **Storage Networking Industry Association (SNIA)-Data Management Forum: The Information Lifecycle Management Maturity Model**  
[www.snia.org/forums/dmf/programs/ilmi/ilm\\_mm/](http://www.snia.org/forums/dmf/programs/ilmi/ilm_mm/)
- **MetaArchive**  
[www.metaarchive.org](http://www.metaarchive.org)
- **Hathi Trust**  
[www.hathitrust.org/](http://www.hathitrust.org/)
- **Disk and Tape Storage Cost Models**  
(Richard L. Moore, Jim D’Aoust, Robert H. McDonald and David Minor)  
[http://chronopolis.sdsc.edu/assets/docs/dt\\_cost.pdf](http://chronopolis.sdsc.edu/assets/docs/dt_cost.pdf)
- **Tiered Fault Tolerance for Long-term Integrity**  
<http://berkeley.intel-research.net/bgchun/bonafide-fast09.pdf>
- **Storage Design for High Capacity and Long Term Storage: Balancing Cost, Complexity and Fault Tolerance**  
<http://www.diglib.org/forums/spring2009/presentations/Farmer.pdf>
- **Keeping Research Data Safe: A Cost Model and Guidance for UK Universities (April 2008)**  
<http://www.jisc.ac.uk/publications/documents/keepingresearchdatasafe.aspx>
- **Image Interchange Framework**  
[http://www.stcatp.org/IIF\\_background.shtml](http://www.stcatp.org/IIF_background.shtml)

## Cloud Computing

- <http://csrc.nist.gov/groups/SNS/cloud-computing/cloud-def-v15.doc>

## MapReduce/Hadoop Framework

- <http://googleblog.blogspot.com/2008/11/sorting-1pb-with-mapreduce.html>
- <http://www.dbms2.com/2009/05/11/facebook-hadoop-and-hive/>
- <http://www.slideshare.net/guest5b1607/text-analytics-summit-2009-roddy-lindsay-social-media-happiness-petabytes-and-lols?type=powerpoint>

## Storage – Possibly Not for Preservation

- <http://blog.backblaze.com/2009/09/01/petabytes-on-a-budget-how-to-build-cheap-cloud-storage/>