UNC
SCHOOL OF INFORMATION AND LIBRARY SCIENCE

Digital Curation and Digital Preservation: An Introduction

JCDL 2007: Tutorial 10

June 19, 2007

Dr. Helen R. Tibbo and Carolyn Hank
School of Information and Library Science
University of North Carolina at Chapel Hill

Outline

- Tutorial Packet
  - Schedule
  - Bibliography
- Introduction
- Key Reports & Issues
- Key Projects & Web Sites
- Developments: Digital Curation Timeline
- Case Study
- Assessment
- Readiness
- Wrap-up

Introduction: Why We Are Here

- Digital preservation (DP) and digital curation (DC) stand as grand opportunities and challenges of the first decade of the 21st century and beyond.

Introduction: Grand Challenges

- “Long-term curation and preservation represent a complex set of challenges, which are exceptionally difficult for data centres and institutions to address individually. They will require a step change in current investment and approaches, and concerted effort on fundamental research, development of shared services, expertise and tools to assist organisations in this work.”
    - JISC Circular 6/03 “An invitation for expressions of interest to establish a new Digital Curation Centre for research into and support of the curation and preservation of digital data and publications”

Introduction: View from the Scientific Community

- The anticipated growth in both the production and repurposing of digital data raises complex issues not only of scale and heterogeneity, but also of stewardship, curation and long-term access.

  - NSF. Cyberinfrastructure Council. “Cyberinfrastructure Vision for 21st Century Discovery.” March 2007. http://www.nsf.gov/pubs/2007/nsf0728/nsf0728.pdf

Introduction: View from the Internet/ Blogger Community

- "WN: The focus of the internet frontier has shifted from the pipes, to search, to community, to blogs, and now video. … What's next?
- Calacanis: Curation. The web and physical world is plagued with abundance -- people need help sorting through all the good and bad stuff out there. The tyranny of choice is causing major psychic pain and frustration for people."

  - "No Stranger to Controversy, Jason Calacanis Starts a New Venture." *Wired* (June 8, 2007). http://www.wired.com/techbiz/people/news/2007/06/calacanis_qa?currentPage=2

Introduction: Definitions and Concepts/ DP

- JISC:
  - "The series of actions and interventions required to ensure continued and reliable access to authentic digital objects for as long as they are deemed to be of value. This encompasses not just technical activities, but also all of the strategic and organisational considerations that relate to the survival and management of digital material."
    - JISC. "Digital Preservation: Continued Access to Authentic Digital Assets." Briefing Paper, November 2006. http://www.jisc.ac.uk/media/documents/publications/digital-pres-bp-v1-04-ab_web.pdf).

Introduction: Definitions and Concepts/ DP

- ALA's Preservation and Reformatting Section (PARS) "Defining Digital Preservation"
  - Short: "Digital preservation combines policies, strategies and actions that ensure access to information in digital formats over time."
  - Medium: "Digital … access to reformatted and born digital content regardless of the challenges of media failure and technological change. The goal of digital preservation is the accurate rendering of authenticated content over time."

- PARS: http://blogs.ala.org/digipres.php

Introduction: Definitions and Concepts/ DP

- ALA's Preservation and Reformatting Section (PARS)
  - Long: "As above plus… Digital preservation policies document an organization's commitment to preserve digital content for future use; specify file formats to be preserved and the level of preservation to be provided; and ensure compliance with standards and best practices for responsible stewardship of digital information. Digital preservation strategies and actions address content creation, integrity and maintenance."

- PARS: http://blogs.ala.org/digipres.php

- So How is Digital Curation Different?

Introduction: Digital Curation

- The active management and preservation of digital resources over the life-cycle of scholarly and scientific interest, and over time for current and future generations of users.

Introduction: Definitions and Concepts/ Digital Curation

- Digital curation involves time-sensitive appraisal by creators and archivists, evolving provision of intellectual access, mid-term preservation including backups and transformations such as migration, and ultimately, for some materials, a commitment to centuries-long archiving.
- Digital curation is stewardship that provides for the reproducibility and re-use of authentic digital data and other digital assets.

Introduction: The DCC on Digital Curation

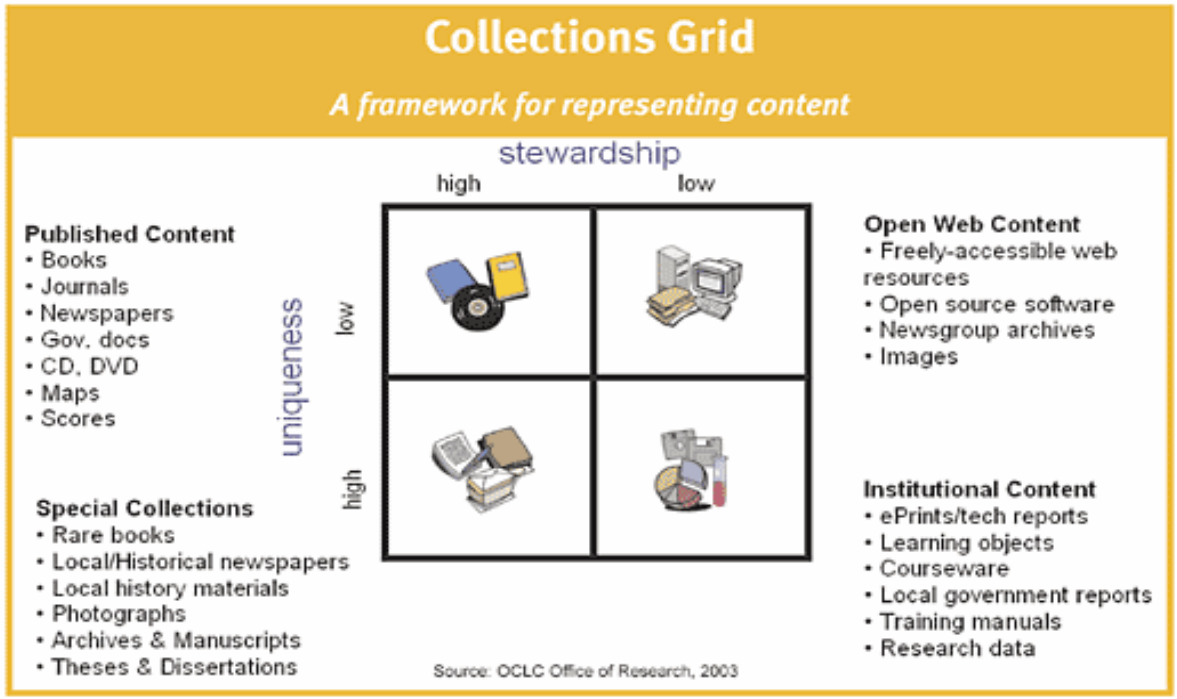
- “Digital curation, broadly interpreted, is about maintaining and adding value to a trusted body of digital information for current and future use.”
- Dynamic as well as static digital objects
- Issues of volatility and scale
- Investigate appraisal and economics
  - “What Is Digital Curation?” http://www.dcc.ac.uk/about/what/

Introduction: JISC on Digital Curation

- "The term digital curation is used in this call for the actions needed to maintain digital research data and other digital materials over their entire life-cycle and over time for current and future generations of users. Implicit in this definition are the processes of digital archiving and preservation but it also includes all the processes needed for good data creation and management, and the capacity to add value to data to generate new sources of information and knowledge."

  JISC Circular 6/03 "An invitation for expressions of interest to establish a new Digital Curation Centre for research into and support of the curation and preservation of digital data and publications" http://www.dcc.ac.uk/about/what/

Introduction: Ecology



OCLC. 2003 Environmental Scan: A Report to the OCLC Membership. Available at: http://www.oclc.org/reports/escan/appendices/collectiongrid.htm

Introduction: Asset Types

- Theses/dissertations/other student papers
- e-Portfolios
- Pre-prints/e-Prints
- Conference Proceedings/Presentations
- Tech reports/working papers
- e-Books
- E-Journal files
- Newspapers
- Datasets
- Databases/Spreadsheets

- University electronic records/publications
- Digital images/audio/moving images
- Digitized Musical Scores
- Exhibitions/performances
- Digital materials: either acquired or created
- PDF files/GIS file/XML files

- Sources: van Westrienen, G. & Lynch, C.A. (2005). http://www.dlib.org/dlib/september05/westrienen/09westrienen.html
  Lynch & Lippincott (2005). http://www.dlib.org/dlib/september05/lynch/09lynch.html

Introduction: Asset Types

- Interview transcripts
- Maps/plans/blueprints
- Software
- Course content/learning objects
- Campus blogs
- Newsletters
- Laboratory protocol
- Manuscripts
- Web pages/sites
- Email

- Sources: van Westrienen, G. & Lynch, C.A. (2005). http://www.dlib.org/dlib/september05/westrienen/09westrienen.html
  Lynch & Lippincott (2005). http://www.dlib.org/dlib/september05/lynch/09lynch.html

Introduction: File Formats

- Thousands of them!
- Select for object creation based on:
  - Open standards
  - Ubiquity
  - Stability
  - Metadata Support
  - Feature Set
  - Interoperability
  - Viability
    - The National Archives. “Selecting File Formats for Long-Term Preservation” 2003. http://www.nationalarchives.gov.uk/documents/selecting_file_formats.pdf

Introduction: File Formats

- Select for preservation based on:
  - Authenticity
  - Processability
  - Presentation
    - The National Archives. “Selecting File Formats for Long-Term Preservation” 2003. http://www.nationalarchives.gov.uk/documents/selecting_file_formats.pdf

Introduction: File Formats

- **PNG** (pronounced 'ping'), the Portable Network Graphics file format, is an open raster image format.
- **JPEG 2000** is an open raster image format described by the ISO/IEC standard 15444, and ITU standard T.800.
- **MrSID**® stands for Multi-resolution Seamless Image Database. Its file format is given the file extension '.sid'. It is designed to compress huge images seamlessly and allow selective delivery and decompression.
- **DjVu**® is a screen/Web format and is more suited to 'mixed documents' (i.e. text and image) than to individual images.
  - Technical Advisory Service for Images. “New Digital File Formats.” http://www.tasi.ac.uk/advice/creating/newfile.html

Introduction: File Format Registries

- **PRONOM**. “PRONOM is a resource for anyone requiring impartial and definitive information about the file formats, software products and other technical components required to support long-term access to electronic records and other digital objects of cultural, historical or business value. “
  - The National Archives. http://www.nationalarchives.gov.uk/pronom/

Introduction: File Format Registries & Tools

- **Global Digital Format Registry (GDFR)**. Harvard University and Andrew Mellon Foundation.
  - http://hul.harvard.edu/gdfr/about.html
- **JHOVE**. (pronounced "jove") The JSTOR/Harvard Object Validation Environment. JSTOR and the Harvard University Library project to develop an extensible framework for format validation.
  - http://hul.harvard.edu/jhove/

Introduction: Key Issues for Digital Curation

- Issues & decisions for long-term DC and DP projects:
  - Creating durable digital objects
  - Appraisal and selection
  - Technologies (e.g., obsolescence, migration, emulation, digital repositories)
  - Risk management
  - Rights management and other legal and ethical issues
  - Digital asset management
  - Metadata (minimum/optimal/practical)
  - Standards
  - File formats
  - Quality control and “trustworthiness”
  - Resource allocation and costing
  - Funding for development and sustainability

Introduction: Threats

- Lack of societal awareness and imperative
- Little national planning and funding
- Lack of institutional policies or planning
- Lack of institutional support and resources
- Lack of local expertise
- Technological obsolescence
- Product/technology development and support

Introduction: Requirements for Digital Curation and Preservation

- Resources – not just more, but new ones
- Trusted and durable digital repositories
- Principles of sound metadata construction
- Use of open standards for file formats and data encoding and
- The promotion of information management literacy.

Key Reports: CPA Archiving Task Force Report

- Garrett, John & Donald Waters. "Preserving Digital Information: Report of the Task Force on Archiving of Digital Information." (The Commission on Preservation and Access and RLG. 1996). http://www.rlg.org/legacy/ftpd/pub/archtf/final-report.pdf

CPA: Challenges

- Technological Obsolescence
- Need for Migration
- Legal Issues
- Institutional Issues
- Need for Deep Infrastructure
- Uncertainty
- People

CPA: Integrity of Digital Information

- Content
- Fixity
- Reference
- Provenance
- Context

CPA: Stakeholder Interests

- Creator
- Archivist
- User
- Society

CPA: Archival Roles & Responsibilities

- Creators
- Archives
- Users

CPA: Migration Strategies

- Change Media
- Change Format
- Incorporate Standards
- Build Migration Paths
- Use Processing Centers

CPA: Managing Costs & Finances

- Cost Modeling
- Obstacles and Prospectives
- Sources of Funding

CPA: Findings

- The first line of defense against loss of valuable digital information rests with the creators, providers and owners of digital information.
- Long-term preservation of digital information on a scale adequate for the demands of future research and scholarship will require a deep infrastructure capable of supporting a distributed system of digital archives.

CPA: Findings

- A critical component of the digital archiving infrastructure is the existence of a sufficient number of trusted organizations capable of storing, migrating and providing access to digital collections.
- A process of certification for digital archives is needed to create an overall climate of trust about the prospects of preserving digital information.

CPA: Findings

- Certified digital archives must have the right and duty to exercise an aggressive rescue function as a fail-safe mechanism for preserving valuable digital information that is in jeopardy of destruction, neglect or abandonment by its current custodian.

CPA: Best Practices & Benchmarking

- Design of systems that facilitate archiving at the creation stage.
- Storage of massive quantities of culturally valuable digital information.

CPA: Best Practices & Benchmarking

- Requirements and standards for describing and managing digital information.
- Migration paths for digital preservation of culturally valuable digital information.

Key Reports: It’s About Time, 2003

- Hedstrom, Margaret et al. “It's About Time: Research Challenges in Digital Archiving and Long-term Preservation.” (Washington, DC: NSF & LOC, 2003).

IAT: Research Challenges

- Technical architectures for archival repositories
- Attributes of archival collections
- Digital archiving tools and technologies and
- Organizational, economic, and policy issues

IAT: Preservation Challenges

- Digital collections are vast, heterogeneous, and growing at a rate that outpaces our ability to manage and preserve them.
- Much more digital content is available and worth preserving; researchers increasingly depend on digital resources and assume that they will be preserved.

IAT: Challenges

- Government, commerce, and personal communications rely on digital information and communications.
- Time is of the essence!
- Threat of interrupted management.

IAT: Digital Archiving Research Agenda

- Attributes of Digital Repositories
- Attributes of Archived Collections
- Tools and Technologies
- Policy and Economic Models

Key Reports: Invest to Save, 2003

- Hedstrom, Margaret & Seamus Ross. "Invest to Save: Report and Recommendations of the NSF-DELOS Working Group on Digital Archiving and Preservation." (Washington, DC: NSF & DELOS, 2003).

ITS: Research Agenda

- Preservation Strategies
- Re-engineering Preservation Processes
- Preservation of Systems and Technology

ITS: Areas of Most Impact

- Self-Contextualizing Objects
- Metadata and the Evolution of Ontologies
- Mechanisms for Preservation of Complex and Dynamic Objects

ITS: Long-Term

- *A period of time long enough for there to be concern about the impacts of changing technologies, including support for new media and data formats, and of a changing user community, on the information being held in a repository*.

ITS: Benefits of Digital Preservation

- Protection and conservation of cultural memory
- Global access to open knowledge and support for cross-disciplinary collaboration.
- Preservation for accountability

ITS: Benefits of Digital Preservation

- Reduction of costs by information re-use
- Foundation of a knowledge economy
- Development of digital libraries

ITS: Principles and Assumptions

- The most distinctive characteristic of digital preservation is its long-term perspective.
- Authenticity and integrity are core requirements.
- Scalability is essential for digital preservation.
- Preservation is a continuous and dynamic process.

ITS: Principles and Assumptions

- Preservation is done within the context of a lifecycle.
- Digital preservation requires shared responsibilities.
- Multiple approaches are needed.
- Digital preservation requires multi-disciplinary research teams.

ITS: Principles and Assumptions

- Digital preservation research does not stand in isolation from practice.
- Preservation is a high-priority research area.

Key Reports: Mind the Gap, 2006

- Digital Preservation Coalition, 2006.
- http://www.dpconline.org/docs/reports/uknamindthegap.pdf
- Culmination of the UK Digital Preservation Needs Assessment (UKNA) study carried out for the Digital Preservation Coalition (DPC) to reveal the extent of the risk of loss or degradation to digital material held in the UK's public and private sectors.

MTG: Findings (1)

- Volume and total value of digital information needed in the ‘information age’ is increasing.
- Organisations often do not have good solutions to the long-term preservation of data.
- Considerable confusion about how to address the problem.

MTG: Findings (2)

- “Very few organisations have some kind of digital preservation strategy in place.”
- “Most organisations are hindered by a lack of clear responsibilities for digital preservation, a problem complicated by the need to involve a range of staff with different skills and the need to involve users at all stages of the information lifecycle.”

MTG: Findings (3)

- “Digital preservation is very much a new discipline which is still being developed, and for which there are few people with the necessary skills. As a result it is a complex activity to undertake and is often perceived as risky.”

MTG: Findings (4)

- “A successful digital preservation solution needs more than just management buy-in, it needs awareness and commitment at all levels throughout the organisation and often collaboration with other organisations.”

MTG: Findings (5)

- “It can be hard to put together a strong cost-benefit justification because the main benefits are often intangible or are “public goods” distributed across time and a range of organisations. The common project-based funding model can hinder digital preservation activities as this often fails to place a value on assets that outlive a project and can be reused elsewhere.”

MTG: Recommendations for Organizations (1)

- Create a long-term proactive information/knowledge management plan.
- Perform regular information audits to measure (and continue to measure) their digital preservation needs, and to ensure that these are being met.
- Consider the long-term value of digital material when putting together plans and budgets.

MTG: Recommendations for Organizations (2)

- Encourage an international ‘market’ for digital preservation tools by linking up with other projects around the world and engaging with software vendors.
- Consider the long-term preservation characteristics of the formats they use. They should work together and with software vendors to encourage the development of open file format standards.

Key Projects & Web Sites(1)

- Arts & Humanities Data Service (AHDS). http://ahds.ac.uk/
- CASPAR: Cultural, Artistic and Scientific knowledge for Preservation, Access and Retrieval. http://www.casparpreserves.eu/.
- Chronopolis: Federated Digital Preservation Across Time and Space. http://globalstor.org/pdf/presentations/Moore-chronopolis.pdf & http://www.sdsc.edu/srb/
- Coalition for Networked Information (CNI). http://www.cni.org
- COnservation OnLine (CoOl) http://palimpsest.stanford.edu/
- Council for Library & Information Resources (CLIR). http://www.clir.org

Key Projects & Web Sites(2)

- DigCCurr. Digital Curation Curriculum Project. http://ils.unc.edu/digccurr
- DigiCULT http://www.digicult.info/pages/index.php
- Digital Curation Center (DCC) http://www.dcc.ac.uk/index
- Digital Preservation Coalition (DPC) http://www.dpconline.org/
- Digital Preservation Europe. (DPE) http://www.digitalpreservationeurope.eu/.
- ERPANET. http://www.erpanet.org

Key Projects & Web Sites(3)

- Library of Congress. Digital Preservation. http://www.digitalpreservation.gov/
- PLANETS (Digital Preservation Research and Technology). http://www.dl-forum.de/englisch/projekte/projekte_eng_2711_ENG_HTML.htm.
- Preserving Access to Digital Information (PADI) http://www.nla.gov.au/padi/
- National Archives & Records Administration (NARA). Electronic Records Archive (ERA) http://www.archives.gov/era/
- Technical Advisory Service for Images (TASI) http://www.tasi.ac.uk/

Developments: Timeline of Activity (1990-95)

- **arXiv**, e-print archives for Physics, Mathematics, Computer Science and Quantitative Biology launched at Los Alamos National Laboratory (now, at Cornell). (**1991)**
  - Source: McKiernan, G. “Scholar-based Innovations in Publishing. Part I: Individual and Institutional Initiatives,” *Library Hi Tech News* 20 no. 2 (March 2003): 19-26.

- Adobe announces the release of **PDF 1.0**, which eventually becomes the standard format for electronic publishing. **(1992)**
  - Source: Timeline: http://www.library.cornell.edu/iris/tutorial/dpm/eng_index.html

Developments: Timeline of Activity (1990-95)

- **National Digital Library Program (NDLP)** launched by Library of Congress. **(1994)**
  - Source: Timeline: http://www.library.cornell.edu/iris/tutorial/dpm/eng_index.html

- **First International Workshop** held in Oxford, England with goal of "the development of materials to support development of draft **ISO** standards effort supporting the long term preservation of digital information obtained from observations of the terrestrial and space environments." **OAIS Reference Model** published seven years later, in 2002. **(1995)**
  - Source: http://nost.gsfc.nasa.gov/isoas/int01/ws.html

Developments: Timeline of Activity (1996)

- **Arts and Humanities Data Service** established.
  - Source: http://ahds.ac.uk/about/ahds-timeline.htm
- **PADI**, or Preserving Access to Digital Information, a National Library of Australia initiative, funded. (Initial call for formation came in 1993).
  - Source: http://www.nla.gov.au/padi/about.html

Developments: Timeline of Activity (1996)

- Brewster Kahle's **Internet Archive** is launched.
  - Source: http://www.archive.org/about/about.php

- A seminal report is published by the CPA and RLG, **"Preserving Digital Information: Report of the Task Force on Archiving of Digital Information"** commissioned by the Commission on Preservation and Access and the Research Libraries Group."
  - Source: ftp://ftp.rlg.org/pub/archtf/final-report.pdf

Developments: Timeline of Activity (1997)

- Humanities Advanced Technology and Information Institute **(HATII)** formed at University of Glasgow.
  - Source: http://www.hatii.arts.gla.ac.uk/hatiihist.html
- The Council on Library and Information Resources **(CLIR)** was created following merger of the Council on Library Resources (CLR) and the Commission on Preservation and Access (CPA).
  - Source: http://www.clir.org/about/history.html
- **CogPrints**, the Cognitive Sciences EPrint Archive, is launched by Stevan Harnad at University of Southampton (UK).
  - Source: McKiernan, G. "Scholar-based Innovations in Publishing. Part I: Individual and Institutional Initiatives," *Library Hi Tech News* 20 no. 2 (March 2003): 19-26.

Developments: Timeline of Activity (1998)

- The Scholarly Publishing and Academic Resources Coalition **(SPARC)** debuts.
  - Source: http://www.arl.org/sparc/about/index.html
- Extensible Markup Language **(XML)** standard is created and Encoded Archival Description **(EAD)** Version 1.0 is introduced.
  - Source: Timeline: http://www.library.cornell.edu/iris/tutorial/dpm/eng_index.html

Developments: Timeline of Activity (1998)

- An **RLG** study finds that 2/3 of archives, libraries, museums, and other repositories had assumed responsibility for digital information, but 42% lacked the capacity to mount, read, and access some of this material.
  - Source: Source: Timeline: http://www.library.cornell.edu/iris/tutorial/dpm/eng_index.html
- PBS broadcasts the CLIR film ***Into the Future: On The Preservation Of Knowledge In The Electronic Age.***
  - Source: Timeline: http://www.library.cornell.edu/iris/tutorial/dpm/eng_index.html

Developments: Timeline of Activity (1999)

- **California Digital Library (CDL)** debuts.
  - Source: http://repositories.cdlib.org/escholarship/more_about.html

- **Resource Description Framework (RDF)** is introduced. RDF is intended to provide metadata interoperability across different communities.
  - Source: Timeline: http://www.library.cornell.edu/iris/tutorial/dpm/eng_index.html

- The project, **International Research on Permanent Authentic Records in Electronic Systems (InterPARES I)** begins.
  - Source: http://www.interpares.org/

Developments: Timeline of Activity (1999)

- Project **CAMiLEON** begins at the Universities of Michigan and Leeds to study the use of emulation as a digital preservation strategy.
  - Source: Timeline: http://www.library.cornell.edu/iris/tutorial/dpm/eng_index.html
- Charles Dollar writes ***Authentic Electronic Records: Strategies for Long-Term Access***.
  - Source: Timeline: http://www.library.cornell.edu/iris/tutorial/dpm/eng_index.html

Developments: Timeline of Activity (2000)

- The digital imaging reference book, including information on digital preservation, ***Moving Theory into Practice,*** is released.
  - Source: Timeline: http://www.library.cornell.edu/iris/tutorial/dpm/eng_index.html

- The **National Digital Information Infrastructure and Preservation Program (NDIIPP)** is created, with up to $100 million in funding allocated by the US Library of Congress.
  - Source: http://www.digitalpreservation.gov/library/about.html

Developments: Timeline of Activity (2000)

- Jeff Rothenberg's **"Using Emulation to Preserve Digital Documents"** is published.
  - Source: Timeline: http://www.library.cornell.edu/iris/tutorial/dpm/eng_index.html

- Cornell project on **Risk Management of Digital Information** offers first assessment of the risks involved in migration for use in cultural institutions.
  - Source: Timeline: http://www.library.cornell.edu/iris/tutorial/dpm/eng_index.html

Developments: Timeline of Activity (2000)

- **National Archives of Australia** implements plan to accept digital records and provide for their continual access over time.
  - Source: Timeline: http://www.library.cornell.edu/iris/tutorial/dpm/eng_index.html
- **EPrints**, an open source, open access repository software created by School of Electronics and Computer Science, University of Southampton, UK, is released.
  - Source: http://en.wikipedia.org/wiki/Eprints and http://www.eprints.org/

Developments: Timeline of Activity (2001)

- **METS 1.1 schema** debuts as an XML standard for encoding descriptive, administrative, and structural metadata within a digital library.
  - Source: Timeline: http://www.library.cornell.edu/iris/tutorial/dpm/eng_index.html

- The OCLC/RLG Working Group on Preservation Metadata releases, ***Preservation Metadata for Digital Objects: A Review of the State of the Art***.
  - Source: Timeline: http://www.library.cornell.edu/iris/tutorial/dpm/eng_index.html

Developments: Timeline of Activity (2001)

- The **Digital Preservation Coalition (DPC)** founded.
  - Source: http://www.dpconline.org/graphics/index.html
- The National Library of Australia releases the PANDORA Digital Archiving System (PANDAS). Version 2 was released in 2002, followed by version 3 in 2004.
  - Source: http://pandora.nla.gov.au/historyachievements.html

Developments: Timeline of Activity (2001)

- **ERPANET: Electronic Resource Preservation and Access Network** project debuted.
  - Source: http://www.erpanet.org/index.php
- **The Budapest Open Access Initiative (BOAI)** forms, from a meeting of the Open Society Institute (OSI).
  - Source: http://www.soros.org/openaccess/

Developments: Timeline of Activity (2002)

- RLG and OCLC publish, ***Trusted Digital Repositories: Attributes and Responsibilities***.
  - Source: http://www.rlg.org/legacy/longterm/repositories.pdf

- First version of **PRONOM**, on online registry of technical information, is released by the National Archives (UK) digital preservation.
  - http://www.nationalarchives.gov.uk/aboutapps/pronom/default.htm

- CCSDS releases the **Reference Model for an Open Archival Information System (OAIS)**: Blue Book CCSDS 650.0-B-1 (2002).
  - Source: http://public.ccsds.org/publications/archive/650x0b1.pdf

Developments: Timeline of Activity (2002)

- **eScholarship Repository** launched by the California Digital Library (CDL).
  - Source: http://repositories.cdlib.org/escholarship/repository_release.pdf
- ***NINCH Guide to Good Practice in the Digital Representation and Management of Cultural Heritage Materials*** is published.
  - Source: http://www.nyu.edu/its/humanities/ninchguide/
- **International Research on Permanent Authentic Records in Electronic Systems (InterPARES II)** begins, following the conclusion of phase I in 2001.
  - Source: http://www.interpares.org/

Developments: Timeline of Activity (2002)

- European Commission IST Support Measure initiative, **DigiCULT: Technology Challenges for Digital Culture**, begins.
  - Source: http://www.digicult.info/pages/index.php
- **DSpace,** an open source digital software platform jointly developed by MIT and Hewlett Packard, is released.
  - Source: http://en.wikipedia.org/wiki/DSpace
- JISC issues call for proposals for their research initiative, **FAIR, Focus On Access to Institutional Repositories,** Programme.
  - http://www.webarchive.org.uk/pan/13734/20060324/www.jisc.ac.uk/indexdc10.html?name=circular_1_02

Developments: Timeline of Activity (2003)

- **Fedora 1.0 (Flexible Extensible Digital Object and Repository Architecture),** a general purpose repository system developed jointly by Cornell University and the University of Virginia, is released.
  - Source: http://www.fedora.info/about/history.shtml

- An OCLC/RLG backed initiative, the **PREMIS, the PREservation Metadata: Implementation Strategies** working group, formed.
  - Source: http://www.oclc.org/research/projects/pmwg/background.htm

Developments: Timeline of Activity (2003)

- The International Internet Preservation Consortium is formed.
  - Source: Timeline: http://www.library.cornell.edu/iris/tutorial/dpm/eng_index.html

Developments: Timeline of Activity (2003)

- **nestor: Network of Expertise in Long-Term Storage of Digital Resources**, debuts. Following this first phase (2003-06), nestor received second phase funding for 2006-09).
  - Source: http://www.langzeitarchivierung.de/modules.php?op=modload&name=PagEd&file=index&page_id=23

- **E-LIS (EPrints on Library and Information Science),** a digital repository, launches.
  - Source: http://eprints.rclis.org/about.html

Developments: Timeline of Activity (2004)

- The International Organization for Standardization publishes: **ISO 15836:2003, Information and Documentation, the Dublin Core Metadata Element Set**.
  - Source: Timeline: http://www.library.cornell.edu/iris/tutorial/dpm/eng_index.html

- Premier of the **Digital Curation Centre (DCC)** by the Joint Information Systems Committee (JISC).
  - Source: http://www.dcc.ac.uk/news/?start=1075593600

- **DAREnet, the network of Digital Academic Repositories**, is launched (The Netherlands).
  - Source: http://www.darenet.nl/en/page/language.view/dare.darenet

Developments: Timeline of Activity (2004)

- **Delos: Network of Excellence on Digital Libraries**, is formed, with funding from the European Commission.
  - Source:http://www.delos.info/index.php?option=com_content&task=view&id=299&Itemid=26

- **Kopal: Co-operative Development of a Long-Term Digital Information Archive** project launches.
  - Source: http://kopal.langzeitarchivierung.de/index_projektverlauf.php.en

- The **European Archive**, a digital library of cultural artifacts in digital form, is incorporated.
  - Source: http://www.europarchive.org/launch-official.php

Developments: Timeline of Activity (2005)

- RLG/NARA Taskforce on Digital Repository Certification release the draft, ***Audit Checklist for Certifying Digital Repositories: Draft for Public Comment*** (2005).
  - Source: http://www.rlg.org/en/pdfs/rlgnara-repositorieschecklist.pdf

- "**A Proposal for a Global Digital Format Registry (GFDR)"** is published by Stephen Abrams and Dale Flecker.
  - Source: https://collaborate.oclc.org/wiki/gdfr/documents.html

Developments: Timeline of Activity (2005)

- **Portico**, an electronic archiving service, is formed. Formerly, Portico operated as the Electronic-Archiving Initiative (2002).
  - Source: http://www.portico.org/about/

- The International Organization for Standardization (ISO) approves the **PDF/Archive (PDF/A)** file format standard.
  - Source: http://www.aiim.org/article-pr.asp?ID=30413

Developments: Timeline of Activity (2005)

- The National Archives and Records Administration (US) awards a $308 million, six year contract to Lockheed Martin to build the **Electronic Records Archives (ERA).**
  - Source: http://www.archives.gov/press/press-releases/2005/nr05-112.html

- The Digital Curation Centre (DCC) releases their first chapter for the **Digital Curation Manual**, Open Source for Digital Curation. Five additional chapters are released through 2006.
  - Source: http://www.dcc.ac.uk/resource/curation-manual/chapters/open-source/

Developments: Timeline of Activity (2005)

- JISC-funded initiative, **Digital Repositories Programme,** starts in support of thirty research and development projects. The initiative carries on as the **Repositories and Preservation Programme**, with funding secured through 2009.
  - Source: http://www.jisc.ac.uk/whatwedo/programmes/programme_digital_repositories.aspx

Developments: Timeline of Activity (2006)

- In June, the nestor Working Group on Trusted Repository Certification publishes version 1 of their criteria for trusted digital repositories; an English-version, ***Catalogue of Criteria for Trusted Digital Repositories (CCTDR)***, is released in December.
  - Source: http://edoc.hu-berlin.de/series/nestor-materialien/8en/PDF/8en.pdf
- The **PLANETS (Digital Preservation Research and Technology)** project debuts.
  - Source: http://www.planets-project.eu/about/
- **CASPAR (Cultural, Artistic and Scientific knowledge for Preservation, Access, and Retrieval)** project starts in April.
  - Source: http://casparpreserves.org

Developments: Timeline of Activity (2006)

- **DigitalPreservationEurope (DPE)** launched website.
  - Source: http://www.digitalpreservationeurope.eu/news-archive/?m=1167609590
- The National Library of New Zealand, the British Library, and Sytec Resources Ltd launch the **Web Curator Tool (WCT)** as an open-source project.
  - Source: http://webcurator.sourceforge.net/
- The Internet Archive's subscription-based service, **Archive-It** debuts.
  - Source: http://news.com.com/8301-10784_3-6067173-7.html
- Digital Preservation Coalition publishes ***Mind the Gap*.**
  - Source: http://www.dpconline.org/docs/reports/uknamindthegap.pdf

Developments: Timeline of Activity (2007)

- DigitalPreservationEurope (DPE) and Digital Curation Centre (DCC) release draft: ***DRAMBORA: Digital Repository Audit Method Based on Risk Assessment***.
  - Source: http://www.repositoryaudit.eu/
- OCLC, CRL, and NARA release, ***Trusted Repositories Audit and Certification: Criteria and Checklist (TRAC)***, a follow up-to the RLG/NARA 2005 draft, ***Audit Checklist for Certifying Digital Repositories.***
  - Source: http://www.crl.edu/PDF/trac.pdf

Developments: Timeline of Activity (2007)

- **DigCCurr2007**, an international symposium on digital curation, is held in Chapel Hill NC, attracting nearly 300 participants.
  - Source: http://www.ils.unc.edu/digccurr2007/
- American Library Association launched **Digital Preservation mailing list.**
  - Source: http://lists.ala.org/wws
- The IMLS-funded **MIRACLE** project releases their institutional repository census (US) findings.
  - Source: http://www.clir.org/pubs/reports/pub140/pub140.pdf

Developments: Terminology

- Digital Repository or Archive:
  - “These two terms are often used interchangeably.
  - OAIS uses ***archive*** when referring to an organization that intends to preserve information for access and use by a Designated Community.
  - ***Digital repository*** often the term used in the DP and DC arenas.

Developments: Terminology

- Institutional Repository
  - "… a university-based institutional repository is a set of services that a university offers to the members of its community for the management and dissemination of digital materials created by the institution and its community members."

Lynch, C. (2002). Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age. ARL Bimonthly Report 226. http://www.arl.org/newsltr/226/ir.html

Developments: Institutional Repositories

- "Most essentially an **organizational commitment** to the stewardship of these digital materials, including **long-term preservation** where appropriate, as well as organization and access or distribution."
- "While operational responsibility for these **services** may reasonably be situated in different organizational units at different universities, an effective IR of necessity represents a collaboration among librarians, information technologists, archives and records mangers, faculty…"

  See Lynch, C. (2002). Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age. ARL Bimonthly Report 226. http://www.arl.org/newsltr/226/ir.html

Developments: Institutional Repositories

- “At any given point in time, an IR will be supported by a set of information technologies, but a key part of the services that comprise an IR is the management of technological changes, and the migration of digital content from one set of technologies to the next as part of the organizational commitment to providing repository services.”
- “An IR is not simply a fixed set of software and hardware.”
  - See Lynch, C. (2002). Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age. ARL Bimonthly Report 226. http://www.arl.org/newsltr/226/ir.html

Developments: Digital Repositories

- OAIS Reference Model:
  - Framework for DP
- Repository systems:
  - In-house (home-grown)
  - Proprietary (hosted or on-site)
  - Open source
    - e.g., DSpace, Fedora, EPrints
- Repository types:
  - Subject-based repositories
  - Institutional repositories (IRs)

Developments: IR Landscape

- CNI Survey (2005) http://www.dlib.org/dlib/september05/lynch/09lynch.html
  - Surveyed 121 US PhD granting inst.; 81 four-year liberal arts colleges
    - Responses from 97 (78.2%) of PhD institutions:
      - 40% report operational IR
      - 88% without IR (52% of total respondents) in planning stages
    - Responses from 35 (43.8%) of liberal arts inst.:
      - 6% (2) report operational IR
      - 21% without IR in planning stages

Developments: IR Landscape, continued

- ARL Survey (Winter 2006)
  http://www.arl.org/bm~doc/spec292web.pdf
  - Surveyed 123 North American academic institutions: 87 respondents (71%)
  - IR Planning and Deployment
    - No Current IR Plans: 19 (22%)
    - IR Planning: 31 (35%)
    - Deployed Operational IR: 37 (43%)

Developments: IR Landscape, continued

- MIRACLE (Census, Fall 2006)
  http://www.clir.org/pubs/reports/pub140/pub140.pdf
  - Surveyed 2,147 North American academic institutions: 446 respondents (20.8% response rate)
  - IR Planning, Piloting, and Deployment
    - No Current IR Plans: 236 (52.9%)
    - IR Planning Only: 92 (20.6%)
    - IR planning and Pilot Testing: 70 (15.7%)
    - Deployed Operational IR: 48 (10.8%)

Developments: DC and DP in Practice

- DSpace at MIT
  http://dspace.mit.edu/

- e-Prints Soton: University of Southampton's Research Repository
  http://eprints.soton.ac.uk

- Fedora: Tuft's Digital Repository (TDR) Program
  http://dca.tufts.edu/tdr/faq

TUFTS Digital Library

TUFTS UNIVERSITY Tufts Digital Repository Program

Assessment: "Trusted" Digital Repositories

- RLG/OCLC Working Group on Digital Archive Attributes in ***Trusted Digital Repositories: Attributes and Responsibilities***. (2002):
  - "…Provide reliable, long-term access to managed digital resources to its designated community, now and in the future," regardless of infrastructure adopted (e.g., local, vendor-mediated, etc.).
  - Must meet basic "expectations."

Assessment: OCLC/RLG (2002)

- Expectations for “Trust:”
  - Meet high-level organizational and curatorial responsibilities and operational responsibilities, including or in addition to:
    - Accepting responsibility for long-term maintenance of content for both contributors and end-users
    - Maintaining an organization system for long-term viability of repository and its content
    - Meeting fiscal demands and requirements for continued sustainability

Assessment: OCLC/RLG (2002)

- More expectations:
  - Designing standards-compliant systems
  - Evaluating systems to assess “trustworthiness”
  - Being explicit and open in meeting contributor and end-user expectations
  - Developing and implementing auditable practices, policies, and services

Assessment: RLG/NARA (2005)

- Taskforce on Digital Repository Certification, **Audit Checklist for Certifying Digital Repositories: Draft for Public Comment** (http://www.rlg.org/en/pdfs/rlgnara-repositorieschecklist.pdf)

- Goal:
  - "Develop criteria to identify digital repositories capable of reliably storing, migrating, and providing access to digital collections."

Assessment: RLG/NARA (2005) continued

- Organized into four audit activities:
  - (1) The Organization
    - Governance and organizational viability
    - Organizational structure and staffing
    - Procedural accountability and policy framework
    - Financial sustainability
    - Contracts, licenses, and liabilities
  - (2) Repository Functions, Processes & Procedures
    - Ingest/acquisition of content
    - Archival storage: Management of archived information
    - Preservation planning, migration and other strategies
    - Data management
    - Access management

Assessment: RLG/NARA (2005) continued

- (3) Designated Community and the Usability of Information
  - Documentation
  - Descriptive metadata appropriate to designated community
  - Use and usability
  - Verifying understandability
- (4) Technologies and Technical Infrastructure
  - System infrastructure
  - Appropriate technologies
  - Security

Assessment: RLG/NARA (2005) continued

- As Resource: Reference to TDR in IR planning (n=36)
  - Use reported by 12 (33%)
  - Non-Use reported 24 (67%)
    - 11 (44%): IR planning per TDR-release
    - 9 (36%): Unaware of TDR's existence
    - 3 (12%): Lack of time, resources, and/or staff
    - 1 (4%): TDR not applicable
    - 1 (4%): Other IR planning priorities
    - 2 (8%): No use personally, but unsure of all IR planning activities.

Hank, C., Tibbo H.R., and Barnes, H. (2007). "Building from Trust: Using the RLG/NARA Audit Checklist for Institutional Repository Planning and Deployment." Paper presented at IS&T's Archiving Conference 2007, May 21-24, 2007, Arlington, VA.

Assessment: TRAC: OCLC/CRL/NARA (2007)

- OCLC, CRL, and NARA, **Trusted Repositories Audit and Certification: Criteria and Checklist** (2007). http://www.crl.edu/PDF/trac.pdf

Assessment: DRAMBORA: DPE/DCC (2007)

- DPE and DCC. **DRAMBORA: Digital Repository Audit Method Based on Risk Assessment.** http://www.repositoryaudit.eu/

Readiness: Working toward Long-term Access

Cornell University's Three-Legged Stool:

- Organizational Commitment
  - Support for persistent access exhibited through policies and procedures
  - e.g., Mission statement, stakeholder agreements, content selection and acquisition, defining "service," deposit agreements, documented DP plan.
  - Existing or new administrative structures
- Technological Infrastructure
  - E.g., OAIS Reference Model
- Resource Allocation
  - Sustainable funding
  - Staffing - Organizational and technical expertise; training

Readiness: Working toward Long-term Access

- Organizational Commitment
  - Support for persistent access exhibited through policies and procedures
  - e.g., Mission statement, stakeholder agreements, content selection and acquisition, defining “service,” deposit agreements, documented DP plan.
  - Existing or new administrative structures

Readiness: Working toward Long-term Access

- Technological Infrastructure
  - Digital object types
  - Existing archival storage use – access copies, master files, and back-up (e.g., online, magnetic tape, removable media – CD, DVD, etc.)

Readiness: Working toward Long-term Access

- Technological Infrastructure
  - Storage procedures (e.g., back-up, off-site, disaster recovery, etc.).
  - Obsolescence (e.g., file formats, storage media, storage drives, hardware and software)
  - Security

Readiness: Working toward … Funding

- Costs
  - Systems (e.g., equipment/hardware/software)
  - Staffing (primary/secondary)
    - E.g., Organizational and technical expertise; training
  - Services (e.g., back-up/maintenance)
  - Others: Supplies, materials, etc.
- Categories
  - Initial/Start-up
  - Sustaining/On-going
  - Contingencies (e.g., damaged equipment)
- Resources
  - E.g., Institutional and/or endowments
  - E.g., Public and/or governmental
  - E.g., Grants (one-time or recurring rewards)

Readiness: Organizational Commitment

- Example from UNC-CH's Digital Curation/Institutional Repository Committee (DC/IRC):
  - Develop a feasible plan that will both serve UNC-CH's curation needs and will place the University in the forefront of such efforts in the Triangle, nationally, and internationally;
  - Design a pilot IR and digital preservation program in partnership with ITS, the University Libraries, and SILS that will support ongoing research;
  - And develop policies, procedures, and long-term digital preservation strategies to benefit the entire campus. This will include strategies to educate the campus community.
- Actual (from MIRACLE findings):
  - Funding for IRs comes or will come from libraries. It is not coming from academic units.

Readiness: Stakeholders and Collaborators

- Recommended:
  - Senior management
  - “ ... Among librarians, information technologists, archives and records manager, faculty, and university administrators and policymakers.” Lynch (2002).
- Actual (from MIRACLE findings):
  - Libraries 40%
  - Archives, central computing, and other academic units: 12%
  - CIO’s office: 6%

Readiness: Collaborators: UNC-CH Example

- University Library, inc.:
  - Library Systems
  - University Archives
  - DocSouth
- Health Sciences Library
- School of Information and Library Science (SILS)
- Renaissance Computing Institute (RENCI)
- Information Technology Services (ITS)
- The Odum Institute
- Ibiblio.org
- UNC Press
- Kenan-Flagler School of Business
- Department of Romance Languages
- Department of Anthropology
- Department of Art

Readiness: “Partner” Roles: UNC-CH Example

- Provost
  - $$$
- RENCI:
  - Hardware acquisition and management
- University Library’s Systems Department:
  - Systems administration
- SILS:
  - Requirements consultation
  - Pilot collections acquisition
  - Student Support
  - Faculty Support

Readiness: Working toward Long-term Access

- Assessment
  - DC and DP needs assessment in relation to institution's teaching, research, and service roles
- Engagement
  - Identification, recruitment, training, evaluation
- Flexible
  - Designing processes and systems that grow
    - E.g., in response to changing needs of community member and digital asset types

Readiness: Resource Recommendations

- 97 “valuable” planning resources provided:
  - 59 (60%) “unique” titles.
- Commonalities in selection (n=59)
  - 8 (14%) selected by 4-6 respondents
  - 6 (10%) selected by 2 respondents
  - 45 (76%) only selected by one participant

Readiness: Resource Recommendations

- Planning
  - CCSDS, Reference Model for an Open Archival Information System (OAIS): Blue Book CCSDS 650.0-B-1 (2002).
    - http://public.ccsds.org/publications/archive/650x0b1.pdf
  - Cornell University Library's Digital Preservation Management (Implementing Short-term Strategies for Long-term Problems) online tutorial.
    - http://www.library.cornell.edu/iris/tutorial/dpm/eng_index.html
  - *The Handbook*, from the DigitalPreservationCoalition.
    - http://www.dpconline.org/graphics/handbook/index.html).

Readiness: Resource Recommendations

- Assessment
  - **DPE and DCC**. DRAMBORA: Digital Repository Audit Method Based on Risk Assessment (2007). http://www.repositoryaudit.eu/
  - **OCLC, CRL, and NARA.** Trusted Repositories Audit and Certification: Criteria and Checklist (2007). http://www.crl.edu/PDF/trac.pdf
  - **RLG/NARA.** Taskforce on Digital Repository Certification, Audit Checklist for Certifying Digital Repositories: Draft for Public Comment (2005). http://www.rlg.org/en/pdfs/rlgnara-repositorieschecklist.pdf

Wrap-Up: Challenges & Future Considerations

- Institutional Repository Context:
  - Lack of consensus in defining an IR
  - Lack of best practices examples
  - Contradictions between conceptualization and actualization
    - Planned: Faculty Deposits
    - Actual: Where are the deposits?
    - Planned: Preservation.
    - Actual: Preservation?

Wrap-Up: Challenges & Future Considerations

- Abby Smith of NDIIPP (in MIRACLE report):
  - “It is one of the paradoxical findings of the survey that there is detectable urgency in the part of libraries to implement institutional repositories, even as early adopters report difficulties in achieving the purposes for which they were built.”

Wrap-Up: Conclusions

- Digital curation involves **all stages** of the lifecycle or continuum
- Digital curation impacts society broadly
- Appraisal is key
- Need to sustain the objects
- Need to sustain the organization
- New vision for libraries & repositories
- New workforce with new skills
- There will be different commitments to persistence.

Wrap-Up: Contact

- Dr. Helen R. Tibbo

  School of Information and Library Science
  201 Manning Hall
  University of North Carolina at Chapel Hill
  Chapel Hill, NC 27599-3360
  Tel: 919.962.8063
  Fax: 919.962.8071
  Email: tibbo@ils.unc.edu

- Carolyn Hank

  School of Information and Library Science
  100 Manning Hall
  University of North Carolina at Chapel Hill
  Chapel Hill, NC 27599-3360
  Tel: 919.259.3191
  Fax: 919.962.8071
  Email: hcarolyn@email.unc.edu