Digital Archiving and Securing Digital Information

Cal Lee
School of Information and Library Science
University of North Carolina, Chapel Hill

North Carolina Electronic Commerce Summit
March 13, 2007
Raleigh, NC

Bits will be Bits (But not for Long)

- Physical media should be stored in appropriate environmental conditions.
- Take care in handling of media.
- Maintain integrity of bit stream through security, checksums, periodic sampling & other validation
- Bit rot & advantages of newer media both call for **periodic refresh & reformatting**.
- Ensuring the **integrity of the bit stream** in such transfers is extremely important.

Digital objects are sets of **instructions for future interaction**

- Digital objects are useless (& don’t even exist) if no one can interact with them
- Interactions depend on numerous technical components
- Only a small part of preservation work is about treating them like physical artifacts.

Obsolescence

"Those who forget the past are condemned to reload it."

- Nick Montfort, July 2000

- All layers undergo change over time, at varying rates.

Rules to Live By

- Recognize & adopt a new definition of “long-term”
- Plan for ongoing care – never trust claims that you won’t have to
- Digital preservation will always involve risks – identify & manage them
- Avoid single points of failure
- Manage & preserve authentic electronic records (don’t just store files)
- Avoid unnecessary software dependencies
- Clearly define rules, roles & responsibilities
- Work with others
- Allocate resources for sustainability

New Conception of “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. This period extends into the indefinite future.”

Source: Reference Model for an Open Archival Information System (OAIS).

Plan for Ongoing Care

- Never trust claims that you won't have to actively manage records over time
- Anticipate changes to the underlying hardware & software (these are inevitable)
- Always maintain the original bitstream
- Periodically sample records to ensure that they're still readable & not corrupted

Identify & Manage Risks

- Organizational (e.g. funding, staffing, mandate, isolation)
- Technological (e.g. lock-in, legacy data, unreliable hardware, need for “digital archeology”)
- Audit yourself – two useful resources released this month
  - Trustworthy Repositories Audit & Certification (TRAC): Criteria and Checklist
  - Digital Repository Audit Method Based on Risk Assessment (DRAMBORA) Toolkit

Avoid Single Points of Failure

- Backup & disaster recovery – periodically tested
- Redundant copies in multiple locations
- Consider storing in more than one file format simultaneously
- Adopt standards that are supported by many independent vendors

More than Just Storing Files

- To be trusted & admissible in the future, records can't just be treated as individual files
- Metadata is just as important as the content
- Required information:
  - Representation information (e.g. how to read & render a particular file format)
  - Provenance – origin & changes over time
  - Reference – identifiers you can count on
  - Fixity (e.g. hash, checksum, digital signature)
  - Context – maintain connections with associated documents

Avoid Unnecessary Software Dependencies

- Good:
  - Standard formats & protocols
  - Publicly available application programming interfaces (APIs)
  - Availability of source code (open-source or code escrow)
- Risky:
  - Encryption
  - Compression
  - Digital rights management
  - Fonts not embedded in files (e.g. consider PDF/A)

Clearly Define Rules, Roles & Responsibilities

- Without these, your records will be at risk internally & suspect externally (by lawyers, judges, corporate customers, citizens)
- Whenever you identify an important requirement, ensure it is someone's job to address it
- A few things to address:
  - Security
  - Access controls
  - Protection of confidential or sensitive data
  - Preservation measures
  - Changes to records & systems
  - Retention & destruction of records
  - Contracts & service agreements with outside parties (including vendors)

Work with Others

- In a digital environment, the lone wolf will die
- Share lessons
- Safety in numbers (minimizes risk)
- Talk to peer organizations in NC and
  - Peer organizations outside NC
  - Department of Secretary of State
  - State Archives
  - Professional associations (e.g. NC Association of Registers of Deeds)
  - Industry associations/consortia
  - Other customers of vendors
  - Standards organizations

Allocate Resources for Sustainability

- Benign neglect will NOT work
- One-time or project funds can be helpful, but you can't depend on them for long-term sustainability
- Build preservation into the cost of doing business
  - Part of annual budget
  - Part of service fees (this is part of the overall cost of an electronic transaction)

Public Key Infrastructure as Authenticity Assurance

- Can address both provenance (who created a record) and fixity (has it been changed since creation)
- Required documentary elements:
  - Private keys
  - Public keys
  - Certificates
  - Certificate revocation lists
  - Algorithms
  - Hash values
  - Signatures as strings of text

Ensuring Authenticity through PKI over Time

- Must maintain the documentary elements I just listed in addition to the digital objects that were signed
- Transformation of bitstream (required for preservation and access over time) breaks the signatures – need to re-sign & maintain trustworthy documentation of previous signatures
- Use of standard formats can minimize this problem

Related Considerations for Notaries (see Electronic Notary Public Act)

- Retain information about specific dates & conditions of the following & associate them with relevant records (be especially attentive to terminations or revocations):
  - Credentials & registration of notary for conducting e-notarization
  - Devices used for signatures

What you Should Ask of Vendors

If I adopt your products/services, what will ensure that I comply with legally mandated records retention requirements (association of records with series, identifying appropriate retention periods, implementing required disposition)?

Prove to me that I can authentically maintain & access my records with someone else's hardware & software in the future.

Show me & explain to me your business continuing & change management plans.

How will you guarantee that my records won't be at risk in 5 or 6 years, when your product line & business model are different?

What has been your approach for migrating data from your previous products?

Put me in contact with customers who went through that migration.

Prove to me that you support relevant open standards.

Show me specifically how you do this.

Provide me contacts outside your company who can vouch for this

Thank You!