soliNET E-cademy

Basic Digital Stewardship

What is Digital Stewardship?

- "a whole range of activities designed to extend the usable life of digital information and protect them from media failure, physical loss, and obsolescence." (Libraries and Archives Canada. "Canada's Digital Information Initiatives." *Toward a Canadian Digital Information Strategy* (2006): 1 August 2006 <http://www.collectionscanada.ca/cdis/012033-302-e.html>.)

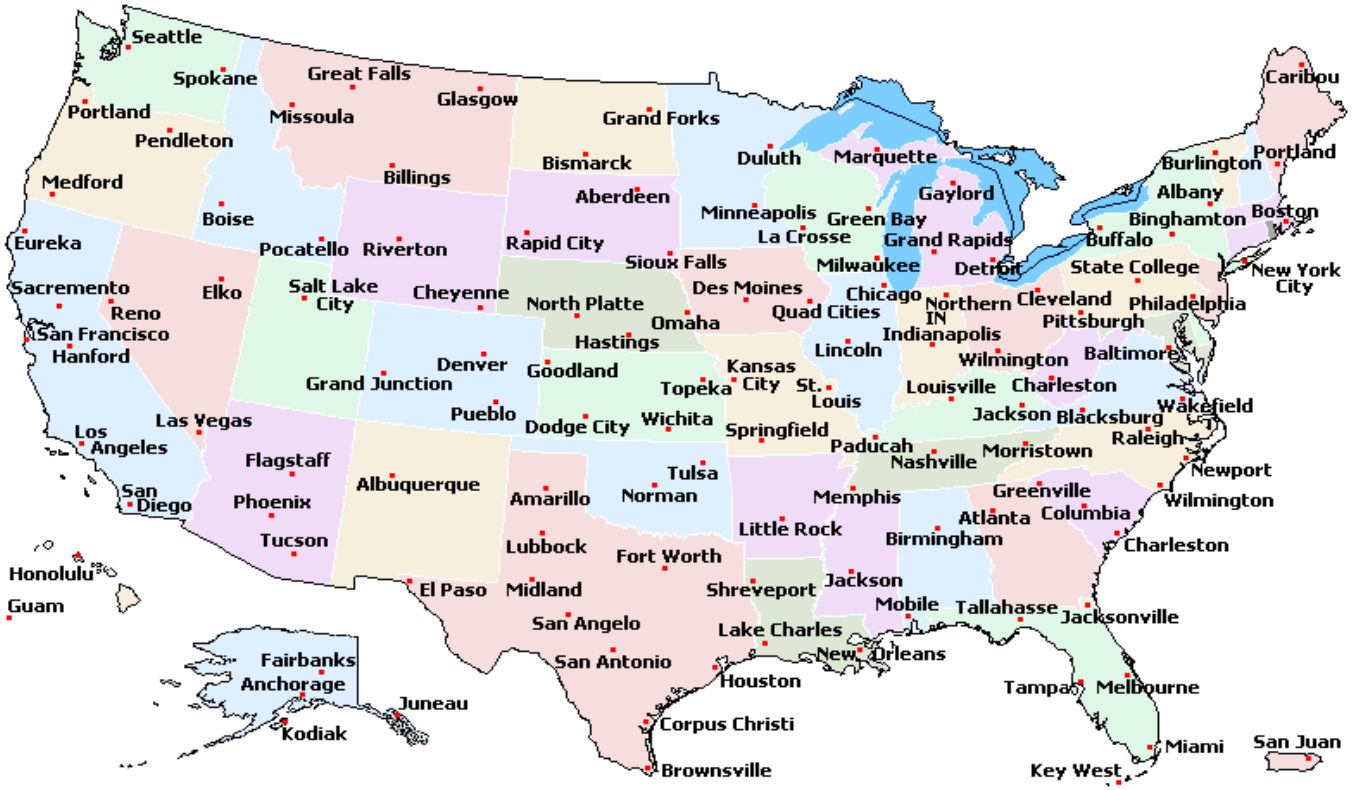
soliNET E-cademy

Using this Software

- Microphone
- Raising your hand
- Green ✓/ Red X
- Laughing / Clapping
- Stepping out
- Text chat
- Feedback
- Audio
- Full Screen
- Exiting

soliNET E-cademy

Show Us Where You Are!



On the toolbar, click on the yellow star, then click on your location!

soliNET E-cademy

Agenda

- Vocabulary
- Strategies for Digital Stewardship
- Choosing a File Format
- Choosing Media
- Storage and Handling
- Metadata

Vocabulary

- Authenticity: the quality of reliability or trustworthiness
- Compression: is a way of encoding a file's data more concisely or efficiently - squeezing the file to reduce file size for storage, transmission, or processing
- Encapsulation: including required access metadata for a digital object inside that object so the two cannot be separated
- Fixity: the state of being fixed or unchanged

Vocabulary

- Lossy/lossless: compression description indicating whether or not information is dropped from files
- Metadata: data that describes (in this case) a digital object, sometimes divided into descriptive, administrative, structural, or other sub-categories
- Normalization: transforming all digital objects of a certain type into the required standard for a particular repository
- OAIS: "Open Archival Information System" – an archival organization of people and systems that has accepted the responsibility to preserve information for a designated community

soliNET E-cademy

Vocabulary

- Proprietary Format: a file format created, managed, and maintained by a commercial organization
- Renderability: data must be viewable by humans and processible by computers
- Understandability: data must be interpretable by humans
- Viability: data must be intact and readable from the storage media

Questions?

soliNET E-cademy

Strategies for Digital Stewardship

- Bitstream Copying
- Refreshing
- Persistent Media
- Standards
- Analog Backup
- Migration
- Data Recovery
- Technology Preservation
- Emulation
- Replication

soliNET E-cademy

Strategies: Bitstream Copying

- Bitstream Copying is the making of an exact duplicate of a digital object.
- Often combined with remote storage so the two copies are not subject to the same disaster (power failure, fire, etc.)
- The minimum required strategy for even the most ephemeral data

Pros:	Cons:
•Easy •Inexpensive	•Doesn't account for loss due to hardware or media failure •Not an adequate long-term strategy for data of significant value

Strategies: Refreshing

- Refreshing is to copy digital information from one long-term storage medium to another of the same type with no change whatsoever in the bitstream.

Pros:	Cons:
•A necessary component to a successful program •Potentially addresses decay and obsolescence of storage media	•Not a complete program itself

Strategies: Persistent Media

- As with cooking, starting with quality, fresh ingredients (such as "gold CDs") is an excellent place to start.

Pros:	Cons:
•Durable media reduces the need for refreshing and diminishes loss from media deterioration	•Has no bearing on other sources of loss (obsolescence, catastrophic destruction, etc.) •Delivers, potentially, a false sense of security

soliNET E-cademy

Strategies: Standards

- Standards do for software what persistent media does for hardware.
- Adhering to well-recognized and broadly implemented standards for encoding and formatting supports those standards and their longevity.

Pros:

•Following industry standards, coincidentally, often means you're also supporting interoperability and potential collaboration.

Cons:

•Industry standards may not meet the specific needs of your collection/institution

Strategies: Analog Backup

- It is possible to transfer some of the content of a digital object back into an analog format (e.g. digital images printed to microfilm)
- Don’t throw out the pre-reformatted materials!

Pros:	Cons:
•Transfers preservation strategy or burden back to media we understand better	•Not suitable for many kinds of digital objects •Sacrifices enhanced operability of digital objects

soliNET E-cademy

Strategies: Migration

- Migration is to convert data from one technology to another while preserving the essential characteristics of the data.
- A preservation specialist at the Rand Corporation compares migration to "preserving a Picasso by repainting it every few years."

Pros:

•Data storage keeps getting cheaper and more reliable
•We're all already doing it—anybody still got a 5.25" floppy drive?

Cons:

•Each transfer provides another opportunity for error and data loss

Strategies: Data Recovery

- Data Recovery is rescuing content from damaged media or hardware
- Usually performed by commercial vendors prepared for broken CDs and other critical damage
- This is an emergency recovery strategy ONLY!

Pros:	Cons:
•Results may be better than a total loss	•Even if the data can be recovered, that doesn't mean that it will be renderable or understandable •Almost no cultural heritage institution will be prepared to do this "in house"

Strategies: Technology Preservation

- Technology Preservation is to preserve the technical environment that runs the system which access your digital collections.

Pros:	Cons:
	•May not be possible to preserve staff understanding of operating the preserved technology •Saving hardware takes more storage than most of us have room for.

Strategies: Emulation

- Emulation combines software and hardware to reproduce the essential characteristics of the a different computer so that media designed for one environment can be used in another one.

Pros:

•You may be able to use files you wouldn't otherwise be able to

Cons:

•Recreating computing environments on different equipment is difficult and not particularly profitable, so the strategy may be unavailable when you need it

Strategies: Replication

- Replication is keeping many copies of the same digital object, preserving copies variously, with the hope that one of them will still be viable when it is needed.

Pros:	Cons:
•Popular/commercial instances show promising systems development (Napster, LimeWire) •It works with popular publishing	•Copies have to be kept in different places and preserved in different ways •Interoperability standards become even more important for multiple participant success

Questions?

Choosing a File Format

- Consider:
  - Wide adoption
  - History of backward compatibility
  - Good metadata support in an open format
  - Built-in error-checking
  - Reasonable upgrade cycles

File Formats: obsolescence

- What causes file format obsolescence?
  - Software upgrades fail to support legacy files
  - The format is superseded by another
  - The format fails to be widely implemented
  - The software that supports the format fails or is withdrawn

File Formats: proprietary v. non-proprietary

- Features may only be available within proprietary formats
- Open formats may include features that are not actually supported by any of the applications available to create or display them
- Open formats take a long time to develop and include input from many interested parties
- Since they are controlled by a commercial company, proprietary formats are vulnerable to being changed or dropped without notice

Choosing Media

- Distinguish your “archive” and “access” collections
  - Archival files are your masters and should be stored off-line and have their access limited
  - Access files are the ones you provide your patrons and most staff and may include a variety of duplicates: thumbnails, screen-sized, higher-resolution, or zoomable versions derived from the same master image

soliNET E-cademy

Choosing Media: stewardship tips

- Follow good storage and handling practices
- Buy quality media
- Track manufacturer and batch numbers to monitor performance trends
- Confirm the fidelity of all media immediately after recording
- Read representative samples on a regular basis

Choosing Media: basic media types

- Magnetic: fixed or removable hard drives, Zip or floppy disks, open-reel and cassette and cartridge tapes
- Optical: CD, CD-RW, DVD
- Solid State: Compact Flash, USB "keychain" drives, flash drives

Choosing Media: hard disks

- Don’t expect more than five years service from any drive
- Make purchasing new drives and refreshing the data part of your initial plan so it can’t be put off when the time comes
- Hard drives are most susceptible to environmental hazards: keep them cool, keep their fans operating, and eliminate dust

Choosing Media: magnetic tape

- High density cartridges (instead of open reel or other designs) are the most reliable, including SDLT, LTO, and AIT
- Per unit, this is the least expensive medium for data storage
- It is uncommon to find backward compatibility beyond one or two generations

Choosing Media: optical

- CDs and DVDs can fail in several areas: dye layer, reflective layer, or surface layer
- Gold is the most reliable reflective layer material
- phthalocynine is the most reliable dye layer material
- The most vulnerable part of the disk is the top layer—often the layer where you find the label!

soliNET E-cademy

Choosing Media: when to use what media

- Offline Storage: CD-R, DAT, DLT
  - Since none of these have been around long enough for anyone to have confident preservation policies for them, write your master files onto two or more of these media and preserve them the best you can
- Online Storage: hard drives
  - Mirror files on redundant disk arrays (RAID), and regularly back up to tape media along with file, Web, mail and other servers for your organization
  - Consider cooperative redundancy storage such as LOCKSS (open source, peer-to-peer software that functions as a persistent access preservation system. Information is delivered via the web, and stored using a sophisticated but easy to use caching system)

Choosing Media: physical threats

- Material instability
- Overuse
- Infrastructure failure (HVAC, plumbing)
- Hardware malfunction
- Sabotage (theft, vandalism)
- Improper storage environment (temperature, dust)
- Natural disaster (fire, earthquake)
- Inadequate hardware maintenance
- Human error

Media Storage

- Among many traditional preservation hazards (infrastructure failure, sabotage, natural disaster, etc.), it is easiest to prevent data losses due to improper storage.
  - Maintain consistent temperature and humidity (68 F and 40% rH)
  - Minimize dust
  - Avoid exposure to magnets
  - No food/drink/smoking
  - Closed metal, electrically grounded cabinets
  - Minimize UV exposure
  - Acclimate media before using
  - Limit access
  - Correctly label

Handling

- Another easy/quick strategy is proper handling of storage media
  - Minimize dust (use lint-free gloves)
  - Clean and dry hands before handling
  - Hold media by edges
  - Avoid flexing (especially optical)
  - Limit access to trained staff
  - Label correctly

Metadata

- What good are multiple copies in multiple formats in multiple places if you don't know what they are, where they are or how to use them?
- Apply similar strategies to preserve metadata on digital objects
- Include preservation-specific fields in "regular" metadata (provenance, context, fixity)
- Metadata should NEVER come last!

Metadata: preservation fields

- Think about including preservation-specific fields in the regular metadata/cataloguing process for digital objects
  - Provenance: for authentication and a documented history of the file's contents
  - Context: why the data was created, how it relates to other data
  - Reference identifiers: ISBN, accession number, etc. to demonstrate the relationship between the digital file and any physical holding you have
  - Technical: to describe the technology environment used to create the digital objects and suggest how the files might be read/used

Questions?

Resources

- The National Digital Information Infrastructure and Preservation Program (http://www.digitalpreservation.gov/)
- Digital Preservation Coalition (http://www.dpconline.org/graphics/)
- European Commission on Preservation and Access (http://www.knaw.nl/ecpa/)
- An Introduction to Digital Preservation (http://www.tasi.ac.uk/advice/delivering/digpres.html)
- Digital Preservation Tutorial (http://www.library.cornell.edu/iris/tutorial/dpm/eng_index.html)
- Statement of Principles for the Preservation of and Long-Term Access to Australian Digital Objects (http://www.nla.gov.au/preserve/digital/princ.html)
- Preserving Access to Digital Information (http://www.nla.gov.au/padi/)
- Care and Handling of CDs and DVDs (http://www.itl.nist.gov/div895/carefordisc/)

Questions

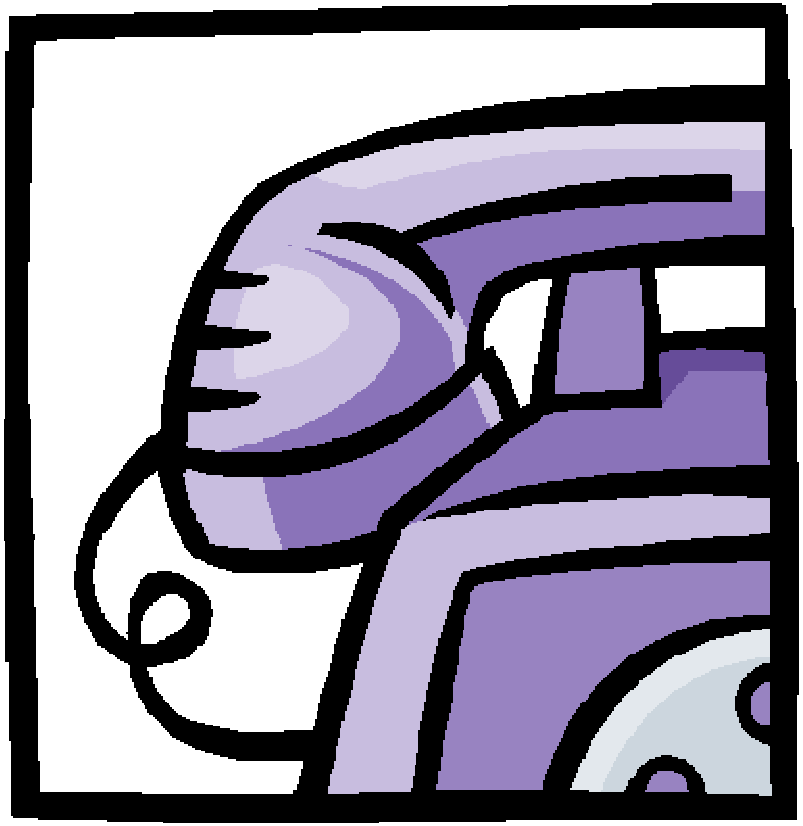
- Digital Stewardship: Courtney S. Danforth (profdanforth@gmail.com)
- Preservation and SOLINET: Kara M. McClurken (kmcclurken@solinet.net)

Evaluation

- Please fill out the evaluation for this course.
  - http://www.solinet.net/survey/classevaluation.htm

soliNET E-cademy

Thank You for Attending!



Questions?

SOLINET Help Desk

1.800.999.8558
helpdesk@solinet.net
8am-5pm Eastern
Monday-Friday