Small steps and lasting impact: making a start with preservation
or
It’s not all NASA

Patricia Sleeman
Digital Archives and Repositories
University of London Computer Centre
(ULCC)
Overview of session

• Not all NASA models
• But it helps – OAIS
• You could be using it already
• Some preservation strategies
• Some useful tools
• How and where do I start!
Overview of session

Flickr Accidentally Wipes Out Account: Five Years And 4,000 Photos Down The Drain

Robin Wauters
10 hours ago
NASA family portrait

VENUS  EARTH  JUPITER
SATURN  URANUS  NEPTUNE

NASA/ P-36089C, 90-HC-380
NASA doesn’t always get it right

• About 20 percent of the data collected for NASA's 1976 Viking Mars landing is completely unreadable and lost forever. With over 1000 people working on the landing alone can you imagine how much they spent to get that 20% of data in the first place?
Consultative Committee for Space Data Systems

RECOMMENDATION FOR SPACE DATA SYSTEM STANDARDS

Reference Model for an Open Archival Information System (OAIS)

CCSDS 650.0-B-1
BLUE BOOK

January 2002

Open Archival Information System (OAIS)
An information package (IP)

• Contains
  – The digital object(s) to be preserved.
  – The metadata required at that point in the system.
  – The Packaging Information which relates 1 and 2.
Information packages in OAIS

• OAIS outlines three types of Information Package:
  – Submission Information Packages (SIPs).
  – Archival Information Packages (AIPs).
  – Dissemination Information Packages (DIPs).
Personal archiving
Identify
Here’s your SIP

Metadata + digital photo.jpeg = SIP
AIP

Ref/name, Context info
Fixity info
+ object = digital photo. TIFF = AIP
Decide how you will preserve your photos over time.

Make sure you look after all the information about your photos and their preservation needs.
Here’s your DIP

Photo = digital object as Jpeg
Descriptive information
Other ways of looking at digital preservation workflow

**Preservation Step-by-Step**

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PRESERVATION PLANNING</strong></td>
<td><strong>RECEIPT &amp; INVENTORY MANAGEMENT</strong></td>
<td><strong>PROCESSING &amp; ARCHIVAL DEPOSIT</strong></td>
<td><strong>MONITORING &amp; MANAGEMENT</strong></td>
<td><strong>CONTENT DELIVERY</strong></td>
</tr>
<tr>
<td>Portico analyzes content and develops a specific preservation</td>
<td>Content is transferred to Portico from the content provider.</td>
<td>Portico ingests content into the archive according to the specific preservation</td>
<td>Portico manages the archive and content on an ongoing basis to ensure</td>
<td>Content is made available to users at participating institutions under defined</td>
</tr>
</tbody>
</table>
DCC Lifecycle model
Some simple preservation strategies

• Build policy foundations
• Develop an internal system
• Use external services
• Learn by doing
• Collaborate
Some simple preservation strategies

- Bit preservation
- Migration
- Emulation
Bit preservation

• Capture information in its original form and focus on maintaining data integrity.
  – Advantages: scalable, practicable; works well (so far).
  – Disadvantage: useful life of data unclear (formats may go obsolete, for example).
Migration

• Transform/normalize data into formats and structures that are optimal for preservation.
  – Advantages: Homogeneous data easier to manage, access; possible to solve preservation issues once and for all if data never has to be transformed again.
  – Disadvantages: Complex initial processes; potential loss of data, functionality; scalability, practicality not proven.
Refreshing is NOT migration

Refresh: Word V2 file → Word V2 file

Modified Refreshing: Word V2 file → Word V2 file

Migration: Word V2 file → PDF file
Emulation

• Use software to mimic behavior of obsolete systems to access and use original data.
  – Advantages: look and feel preserved; no need to process original files.
  – Disadvantages: development of emulators can be costly and needs to account for application and operating system, both of which can have many variations; always some uncertainty if the emulation is completely correct.
You are not alone

• Collaboration is key to digital preservation
  – Use your contacts and identify expertise which exists elsewhere which you may need
  – Funding bids
  – Examples of collaboration include
    • NDAD/TNA project
    • DPTP training project
Functions which tools are needed for

- Preservation planning including policy and strategy planning, risk assessment, software decisions, implementing standards
- Data objects are appropriately ingested, archivally stored and managed
- Administrative procedures are in place for the overall operation of the archive
- Data objects continue to be accessible to those who need to use them over time
Planning: where and how do I start?

Credits: Nancy McGovern & Ann Kenny, Cornell University
AIDA

- AIDA self assessment toolkit
  - Measures capacity
  - Readiness
  - Capability

http://aida.jiscinvolve.org/wp/
Some tools to look at

- **Bagit** - the Library of Congress tool
- **DROID** - National Archives' tool for file format identification
- **Jhove** - developed by JSTOR and Harvard for object identification, validation and metadata extraction
- **NLNZ metadata extraction tool** - extracts basic metadata from some popular formats
Tools – good sites to visit

Comprehensive advice sites

– Digital Preservation Coalition
  [http://www.dpconline.org](http://www.dpconline.org)

– PADI

– Library of Congress
Training

• Training opportunities
  – www.dptp.org
In the meantime…small steps

• Decide how you will preserve your objects over time.
• Think about the best preservation format.
• Make at least two copies of your object—more copies are better.
• Store copies in different locations that are as physically far apart as practical. If disaster strikes one location, your photographs in the other place should be safe.
• Put a copy of the inventory in a secure location.
• Check your objects at least once a year to make sure you can read them.
• Create new media copies every five years or when necessary to avoid data loss.
In the meantime

**Good practise**

- Use open, non-proprietary data formats.
- Try to make sure your metadata in conformance with emerging standards and documentation aimed at facilitating future use and future management of the resource.
What else?

• Look at successes
  Gloucestershire archives:
  http://futurearchives.blogspot.com/2010/03/scat-gloucestershire-archives.html

• Look at failures
  – Both are learning opportunities

• Look at DPC case notes series
  http://www.dpconline.org/advice/case-notes

• Share your successes/failures
And remember

• You are probably doing a lot of it already…
Thank you

Patricia Sleeman

P.sleeman@ulcc.ac.uk

- http://www.dptp.org
- http://dablog.ulcc.ac.uk/
- http://www.ulcc.ac.uk