Digital Preservation

Sharon McMeekin
Head of Training and Skills
Digital Preservation Coalition
Part One:
Why should we care about Digital Preservation?
Growing at an Exponential Rate

The Great Recession hasn’t slowed the tremendous growth of the Digital Universe. In 2016, the volume of digital information created and duplicated in a year will reach 1.2 zettabytes, according to new data from IDC.

Figure 1: The Digital Universe 2009–2020
Growing by a Factor of 44

2020
35 ZB*

Data is the New Oil?

Cash Value: $6.3bn
Share Valuation: $104bn
Intangible assets: $97.7bn
(10 dollars every 3 seconds non stop for 950 years)
Pieces of data: 2.1trillion
Price per datum: $0.05
“...The crux of data’s worth is its seemingly unlimited potential for re-use: it’s option value. ...most of data’s value lies in its use not in its possession”
(Mayer-Schonberger and Cukier 2015)
It Needs to be Useable (2)

‘It’s not like finding a needle in a haystack: it’s like finding a needle in Germany.’

(Paul Miller, DPC Briefing Day on Linked Open Data, July 2013)

It Can Disappear Quickly
Corporate Abandonment

Nook
GoogleCode
GeoCities
GoogleWave
Knol
Friends ReUnited
Yahoo Mail

Blipfoto
MySpaceBlogs
Yahoo 360
del.icio.us
MyBlogLog
BeBo
Classic

www.digitalbevaring.dk
People are a Big Threat

Accidents Happen
but also...
People With Things to Hide....
• US Environmental Data
• Hillsborough Enquiry
• Leave.EU
Why We Preserve....

- Legal and Regulatory Compliance
- Increased Efficiency
- New Revenue Streams
- Improving Health
- Protecting the Environment
- Enabling Research
- Documenting Cultural Heritage
- Ensuring Transparency and Accountability
Part Two:
So, what’s the problem?
Traditional Media

- Robust
- Tangible
- Independently understandable
- Well-developed approaches to preservation
- Experienced in assigning value

- Traditional objects are generally quite robust
- They are tangible, we can hold them in our hands
- Are generally independently understandable (if you speak the language they are written in.....)
- We are quite experienced in understanding their worth and assigning value to such objects
Digital Information

• Ephemeral
• Need tech to interpret
• Obsolescence
• How to estimate value?
• New skills and solutions required
• But also new opportunities!

• Digital objects are ephemeral by their very nature
• They very susceptible to obsolescence as they are entirely dependent on the media they are stored on, the accessibility of their file format and often require documentation to use and understand them
• Managing issues such as rights can also be much more difficult, from protecting copyright to ensuring personal data is protected
• They require us to gain new skills to care for them, or for us to work with new groups of colleagues with different skills groups (particularly IT specialists)
• But they do also bring a whole host of new benefits, in particular the ability to make content accessible to users.
<table>
<thead>
<tr>
<th>Risks to Digital Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Media failure</td>
</tr>
<tr>
<td>Media obsolescence</td>
</tr>
<tr>
<td>Software failure</td>
</tr>
<tr>
<td>Software obsolescence</td>
</tr>
<tr>
<td>Inaccessible file format</td>
</tr>
<tr>
<td>Format migration errors</td>
</tr>
<tr>
<td>Loss of media</td>
</tr>
<tr>
<td>Disaster</td>
</tr>
<tr>
<td>Broken dependency</td>
</tr>
<tr>
<td>Failure of systems integration</td>
</tr>
<tr>
<td>Data Theft</td>
</tr>
</tbody>
</table>
What’s the Problem?

Digital data (images, documents etc.) have value and create opportunities
...but...
Access depends on software, hardware and people
...and...
Technology and people change, creating barriers to reuse
...therefore...
We need to actively manage data to protect and create opportunities
What is Digital Preservation?

“the series of managed activities necessary to ensure continued access to digital materials for as long as necessary”

(Digital Preservation Handbook)

Basic definition of DP

‘Managed activities’ important, could be rephrased as managing the risks that threaten digital objects
Part Three:
Seems like Digital Preservation is hard.....
AKA – Stuff I wish I’d known about/had access to.....
Where Do I Start?
Start with Simple Steps

- Risk Management
- Maturity Modelling
- Bit-Level Preservation
- Digital Asset Register
- Policy


DPCMM - http://www.securelyrooted.com/dpcmm/
Models: Three Legged Stool

http://dpworkshop.org/dpm-eng/conclusion.html
Models: OAIS
Other Standards

• Metadata – METS and PREMIS
• Information Packages – E-ARK Project
• Audit and Certification – Core Trust Seal, ISO 16363...
• Data Security
• File Formats
Vendors/Service Providers

• Arkivum
• Artefactual
• Ex Libris
• FormPipe
• Hanzo
• Keep Solutions
• Libnova
• Mirrorweb
• Preservica
• TownsWeb
Isn’t It Expensive?

www.digitalbevaring.dk
Not in Comparison to the Costs of Looking After Physical Collections!

Setup:
Tens of thousands?

Setup:
Tens of millions?
Is Digital Preservation Expensive?

No!

• But it is an unfunded mandate
• Don’t throw money at it
• Get the mandate properly incorporated
Do We Have the Skills?

www.digitalbevaring.dk
A New Profession?

- Skills gaps
- Increasing recruitment
- Collaborative field
- Traditional Skills Relevant
- Be Adaptable
- Well paid!
The People are Great and I Get to Travel!

- UK
- Ireland
- The Netherlands
- Germany
- Belgium
- Italy
- Austria
- USA
- Hungary
- Portugal
- Slovakia
And Finally:
A Little on the DPC
Digital Preservation Coalition

*We exist to “secure our digital legacy”*

80+ Members and 6 Supporters

http://www.dpconline.org
We Do 6 Things...

- Advocacy
- Community Engagement
- Workforce Development
- Capacity Building
- Good Practice and Standards
- Management and Governance
Advocacy

- BitList
- Digital Preservation Awards
- Advocacy Tools
- Advocacy Support
Community Engagement

• Webinars
• Blog
• Networking
  • Unconference
• Email List
• World Digital Preservation Day
Workforce Development

- Training
  - Getting Started
  - Making Progress
- Leadership Programme
- Grants
- Student Conference
- Labour Market
- Intelligence
- Curriculum Development
- DP Handbook
Digital Preservation Handbook

Welcome to the revised 2nd edition of the Digital Preservation Handbook. A key knowledge base for digital preservation, peer-reviewed and freely accessible to all.

*Overall the improvements to the Handbook make it, in my opinion, one of the more useful and flexible tools for identifying, understanding and getting to grips with practical approaches to the varying challenges of digital preservation. It uses accessible language, clear terminology and provides useful links to case studies and further reading which is of..."
Capacity Building

- Technology Watch Reports
- Briefing Days
- Digital Preservation Anonymous
- Research Projects
Good Practice and Standards

- Standards Development
  - OAIS Review
- Understanding member needs
- Talking to vendors
- Certification and accreditation
Good Governance

- Member owned and member led
- Transparent
- Accountable
- Responsive
www.dpconline.org
sharon@dpconline.org
@SharonMcMeekin