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

‘Digital Universe’ Nears A Zettabyte
May 4th, 2010: Rich Miller

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

2020
35 ZB*

*Zettabyte = 1 trillion gigabytes

Source: IDC Digital Universe Study, sponsored by EMC, May 2010

The Great Recession hasn’t slowed the breakneck growth of the Digital Universe. In 2010 the volume of digital information created and duplicated in a year will reach 1.2 zettabytes, according to new data from IDC.
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-Schönberger and Cukier 2015)
‘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  Blipfoto
GoogleCode  MySpaceBlo
GeoCities  Yahoo 360
GoogleWave  del.icio.us
Knol  MyBlogLog
Friends ReUnited  BeBo
Yahoo Mail  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
Digital Information

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

- Media failure
- Media obsolescence
- Software failure
- Software obsolescence
- Inaccessible file format
- Format migration errors
- Loss of media
- Disaster
- Broken dependency
- Failure of systems integration
- Data Theft
- Virus or malware
- Poor documentation
- Poor rights management
- Lost encryption key
- Poor version control
- Inadvertent deletion
- Malicious deletion
- Poor file-naming
- Corruption
- Link rot
- ...
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)
Part Three:
Seems like Digital Preservation is hard.....
AKA – Stuff I wish I’d known about/had access to.....
Where Do I Start?

www.digitalbevaring.dk
Start with Simple Steps

- Risk Management
- Maturity Modelling
- Bit-Level Preservation
- Digital Asset Register
- Policy
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?
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?
A New Profession?

• Skills gaps
• Increasing recruitment
• Collaborative field
• Traditional Skills Relevant
• Be Adaptable
• Well paid!

www.digitalbevaring.dk
DigCurV Framework

- Understands creation, management and monitoring of project plans
- Understands project management concepts and techniques
- Understands data management requirements
- Is able to undertake project management activities and innovative practices
- Is able to deal with data curation challenges through structured planning
- Is able to produce relevant information to support decision-making

- Understands audit of curation functions
- Understands certification of repositories or programmes
- Is able to maintain documentation in preparation for audit process

- Is able to apply risk management practices, techniques and standards to digital curation activity within institutional risk management context
- Is able to assess, analyse, monitor, communicate risks

- Understands social and ethical responsibility in digital curation
- Understands energy consumption, carbon footprint of digital curation activity
- Is able to adhere to principles of ethical conduct

- Understands institutional’s legal culpabilities in digital curation activity
- Understands institutional regulatory frameworks within which digital repositories operate
- Is able to apply appropriate actions to curation workflow to ensure compliance with legal and policy frameworks and relevant standards
- Is able to select and apply validation techniques against policy infringement

- Understands legal frameworks in which curation is taking place
- Understands domain policies and standards for management and preservation of digital objects
- Is able to articulate knowledge and definitions
- Is able to select and apply digital curation and preservation techniques
- Understands scope of own role within institutional context

- Is able to articulate benefits and long-term value of collections
- Understands information-technology definitions and skills
- Is able to plan application of selection/appraisal criteria to collections
- Is able to conduct user needs analysis
- Is able to continuously monitor and evaluate digital curation technologies

- Understands information-seeking strategies, access technology and sharing behaviours
- Understands how to support information access and sharing
- Is able to deploy appropriate information-seeking strategies
- Understands key metadata standards for sector/subject
- Is able to select and apply metadata standards
- Understands relationship between appropriate controlled vocabularies and metadata standards

- Understands data structures and types
- Understands file types, applications and systems
- Understands databases types and structures
- Is able to execute analysis of and forensic procedures on digital objects
- Is able to apply responsibility, accountability and good practice in digital curation
- Is able to identify malpractice
And Finally:
A Little on the DPC
Digital Preservation Coalition

We exist to “secure our digital legacy”

80+ Members and 6 Supporters
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
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 approachable language, clear terminology and provides useful links to case studies and further reading which will be 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