Digital Preservation

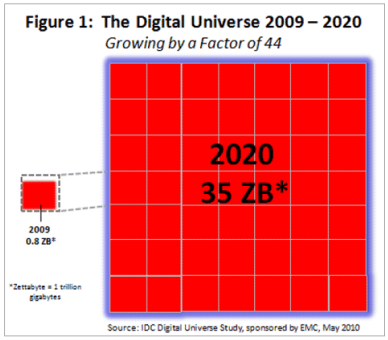
Sharon McMeekin
Head of Training and Skills
Digital Preservation Coalition

Part One:
Why should we care about Digital Preservation?

Growing at an Exponential Rate

'Digital Universe' Nears A Zettabyte

May 4th, 2010 : Rich Miller



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

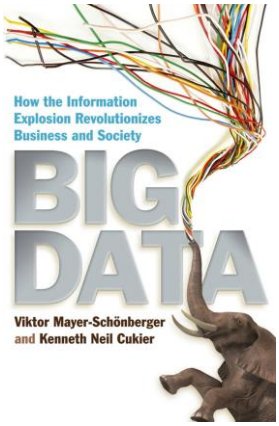
Intangible assets: $97.7bn

(10 dollars every 3 seconds non stop for 950 years)

Pieces of data: 2.1trillion

Price per datum: $0.05

It Needs to be Useable (1)



"…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)

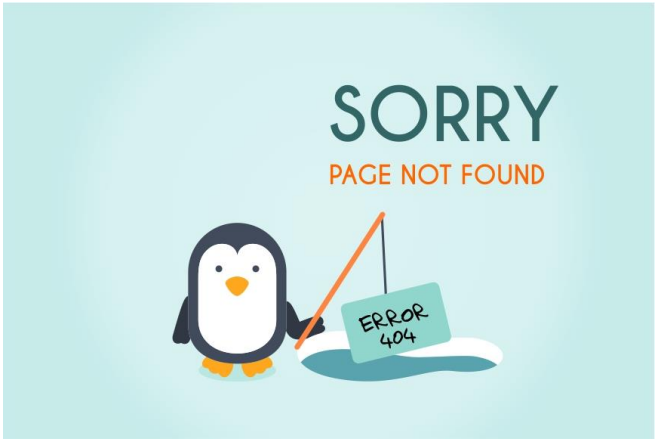
It Needs to be Useable (2)

'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)

G. Bembridge 2012 https://www.flickr.com/photos/tipsfortravellers/

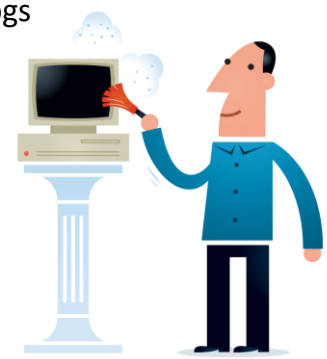
It Can Disappear Quickly



Corporate Abandonment

Nook
GoogleCode
GeoCities
GoogleWave
Knol
Friends ReUnited
Yahoo Mail

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

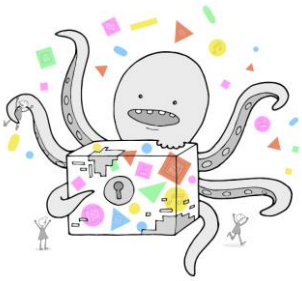


www.digitalbevaring.dk

People are a Big Threat

Accidents Happen
but also…
People With Things to Hide….

- US Environmental Data
- Hillsborough Enquiry
- Leave.EU

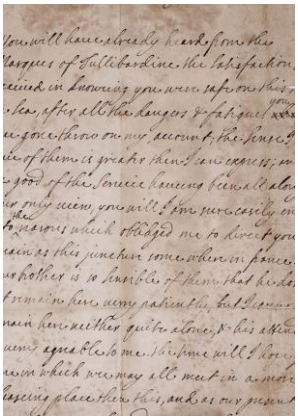


Why We Preserve….

- 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.

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)

Illustration by Jørgen Stamp digitalbevaring.dk CC BY 2.5 Denmark

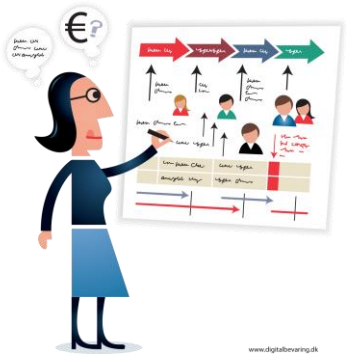
Basic definition of DP

'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

NDSA Levels of Digital Preservation - https://ndsa.org//activities/levels-of-digital-preservation/

DPCMM - http://www.securelyrooted.com/dpcmm/

Models: Three Legged Stool

http://dpworkshop.org/dpm-eng/conclusion.html

Models: OAIS

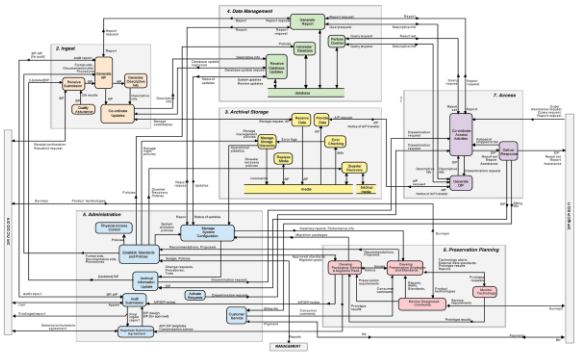


Figure F-1: Composite of Functional Entities

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

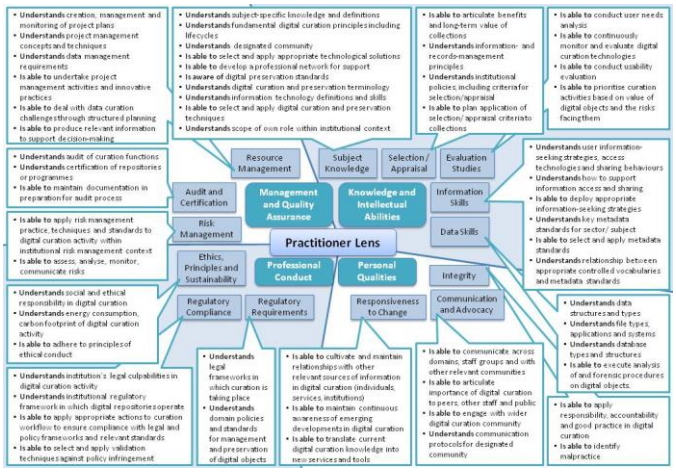
Do We Have the Skills?

www.digitalbevaring.dk

A New Profession?

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

DigCurV Framework



DigitalPreservationCoalition

The People are Great and I Get to Travel!

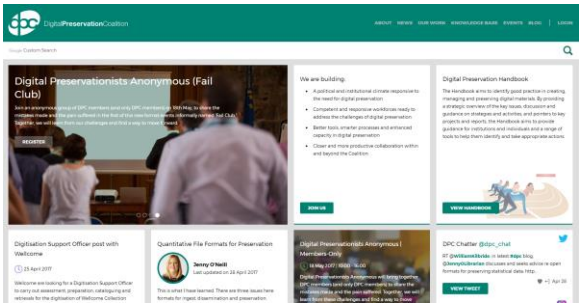
- UK
- Ireland
- The Netherlands
- Germany
- Belgium
- Italy
- Austria
- USA
- Hungary
- Portugal
- Slovakia

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

LLOYDS BANKING GROUP
HSBC
THE UNIVERSITY OF MELBOURNE
Jisc
Data Archiving and Networked Services
DANS
100
BRITISH LIBRARY
An Chartlann Náisiúnta
National Archives
Trinity College Library Dublin
Library and Archives Canada
EUROPEAN CENTRAL BANK
EUROSYSTEM
BANK OF ENGLAND
BACARDI
University of Salford MANCHESTER
The Open University
Historic England
dri
National Records of Scotland
LOCKSS
RESEARCH COUNCILS UK
the national archives
ULCC
NATIONAL LIBRARY of IRELAND
University of Glasgow
UNIVERSITY OF CAMBRIDGE
UKLA
RIBA
ACADEMIC PRESERVATION TRUST
Chester Beatty
Unilever
NDA
Nuclear Decommissioning Authority
PORTICO
UNIVERSITY OF LEEDS
UNIVERSITY OF BIRMINGHAM
LGC NLW
ual
LONDON SCHOOL of HYGIENE & TROPICAL MEDICINE
LSE
THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE
DPN
THE DIGITAL PRESERVATION NETWORK
BBC
DCC
ARA
Archives & Records Association UK & Ireland
UNIVERSITY OF BATH
THE UNIVERSITY of York
PRONI
Public Record Office of Northern Ireland
UNITED NATIONS
Mechanism for International Criminal Tribunals
nls
National Library of Scotland
MANCHESTER 1824
UCL
ads
ARCHAEOLOGY DATA SERVICE
CERN
University of Kent
Llywodraeth Cymru
Welsh Government
US
UNIVERSITY OF SUSSEX
KING'S College LONDON
Durham University
THE UNIVERSITY OF WARWICK
Lancaster University
Bibliothèque nationale de Luxembourg
Nationalbibliothéik
UNIVERSITY OF Hull
University of Portsmouth
UCD
ABERYSTWYTH UNIVERSITY
UK Data Archive
Publications Office of the European Union
TATE
The British Museum
Your Logo Here

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

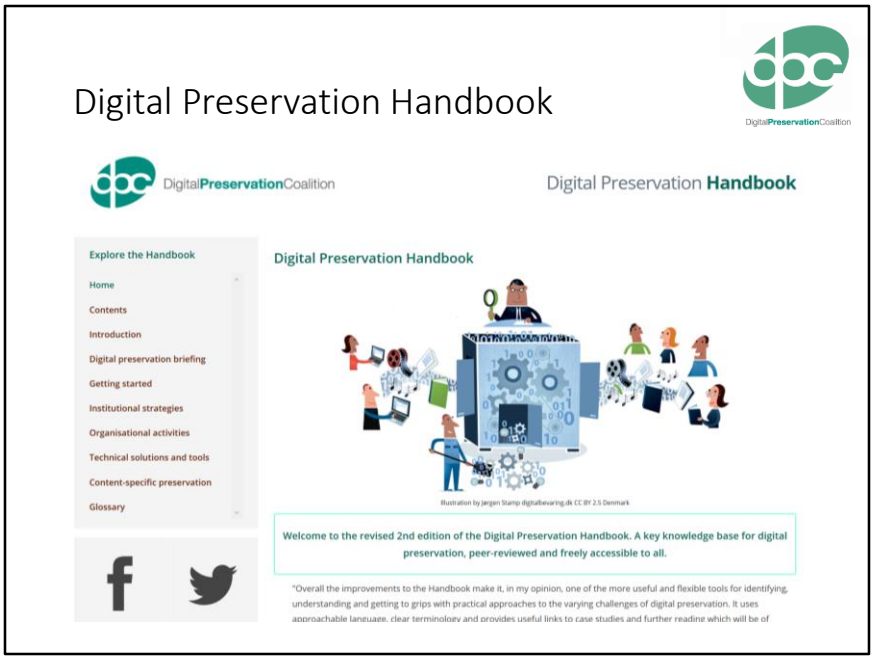
Community Engagement

- 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



http://dpconline.org/handbook

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