

Digital Resources: To Infinity and Beyond?

Curating & Preserving Digital Library Resources

Maureen Pennock

Digital Curation Centre, UKOLN, University of Bath



This work is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 2.5 UK: Scotland License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-sa/2.5/scotland/>; or, (b) send a letter to Creative Commons, 543 Howard Street, 5th Floor, San Francisco, California, 94105, USA.

Funded by:

JISC



Today's talk

- Background
 - What's the problem?
 - What can you do about it?
- Practical tips and helpful initiatives

From Analogue to Digital



from
this...

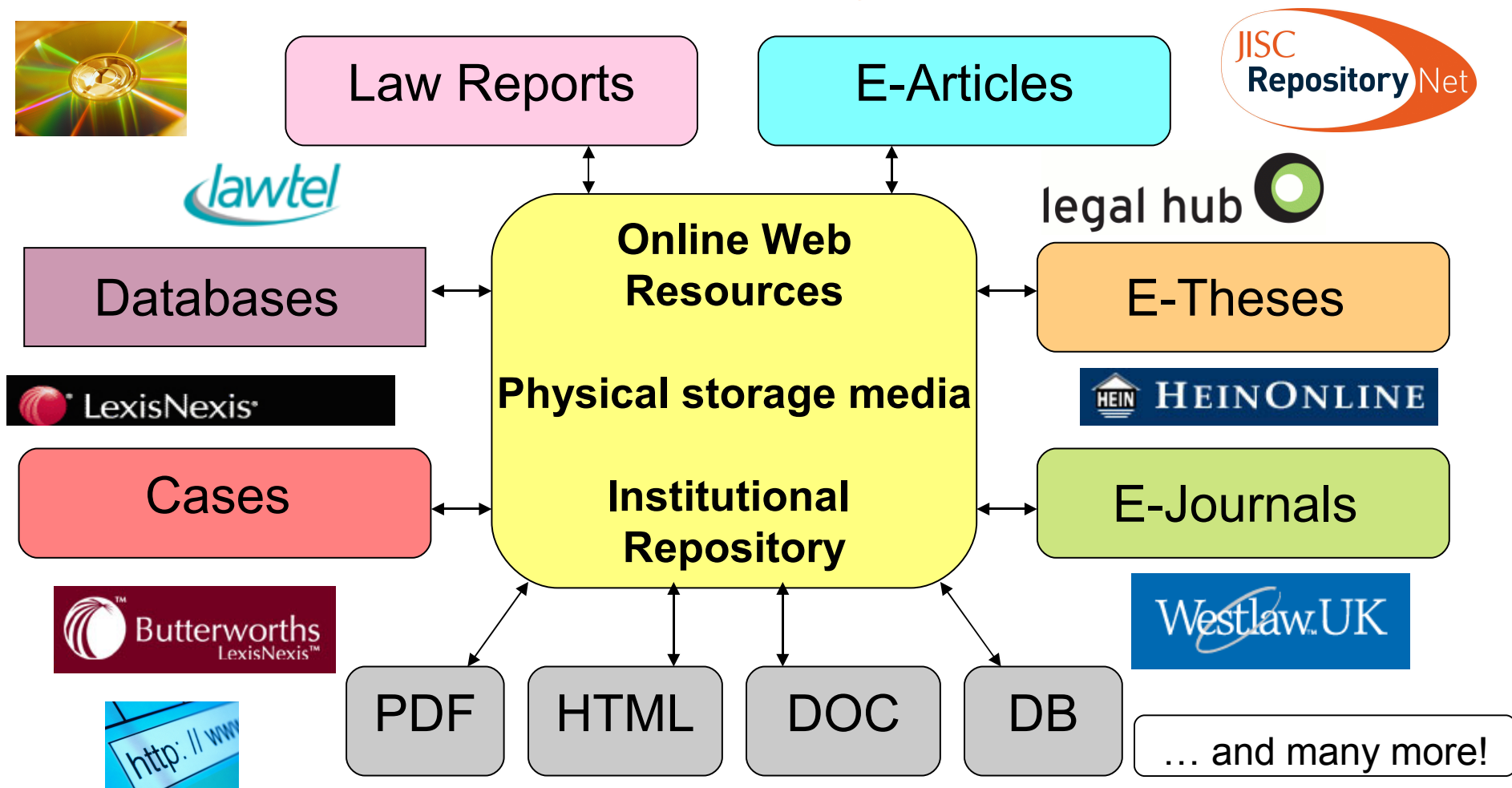


... to this



- Digital resources now commonplace
- Users expect digital resources!
- 'Some' print resources being abandoned

Resource types



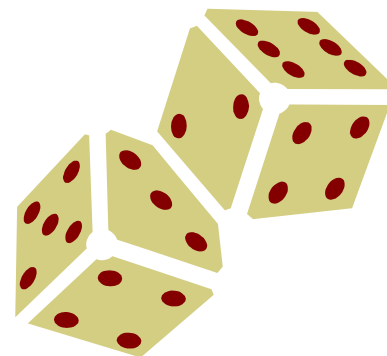
The world at your fingertips!

- But digital materials are fragile

“Digital Information Lasts Forever - Or Five Years, Whichever Comes First”

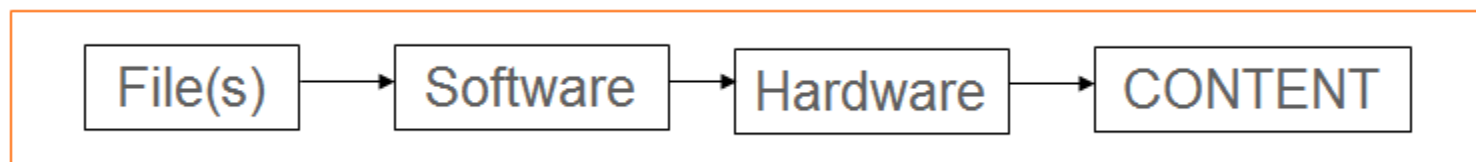
Jeff Rothenberg, 1995

- Risks from:
 - Technological obsolescence
 - Organisational obsolescence
 - Inaction
 - Corruption
 - Malicious intent
 - Users!



Technological challenge

- Digital content is contextually dependent



- Change in context affects availability of content
- Context invariably changes over time
 - Hardware
 - Software
 - Media
- Action must be taken to prevent loss of data



Example 1 – BBC Domesday

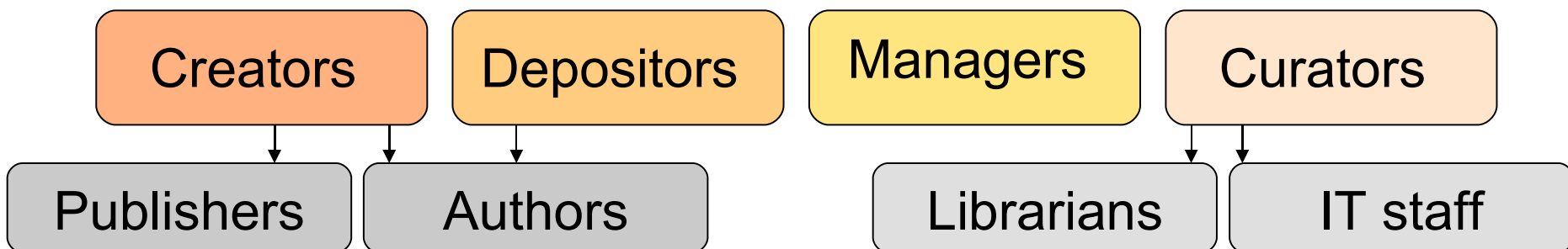
- Incredibly valuable historical resource
 - Text
 - Images
 - Early VR
 - Multimedia
 - Data
- Obsolete by early 2000
- Significantly outlived by the analogue version!
- ‘Heroic rescue’ needed



Photo from CAMILEON website

Organisational challenge

- Changing nature of curatorial responsibilities
- Responsibilities unclear
- Organisational and cultural infrastructure not geared towards digital preservation
- Low awareness (and interest)
- Ongoing challenge involving a range of stakeholders:



Example 2 – Websites

- Average lifespan: 44 days (same as a housefly!)
- Constantly changing
- Neglect by owners
- Domain name expiry
- External link rot
- Internal upgrades leading to errors



Other challenges

- Lack of funding – ongoing commitment; who will pay?
- Absence of preservation tools & services
- Security
- Storage models
- Legal issues:
 - Copyright both an internal and external problem
 - Meeting legal obligations
- User access: How?
- **Failure to address these challenges inevitably results in loss from various quarters**

Other examples

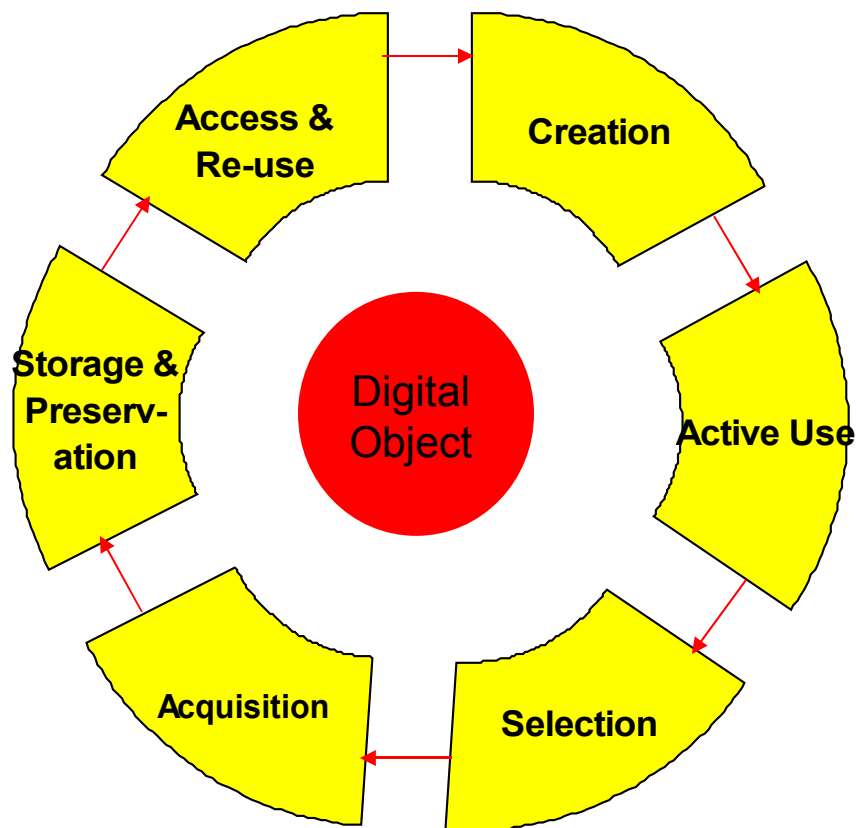
- A collection of 8" floppies with no labels and no equipment to read them
- A computer running old software and files – but with no instruction manual or index!
- Modern CD ROMs containing old MS v1.00 Word files.
- An obsolete/overwritten/ill-maintained database
- A migrated and subsequently corrupt database
- Storage media long ago ‘stored’ on a dusty shelf
- Old games (remember Chuckie Egg?!)
- And what of the future... PDF's?

What Can You Do?

Digital Curation

- Digital Curation, broadly interpreted, is about **maintaining** and **adding value** to a trusted body of digital information for **current and future use**
- The active management and appraisal of data over the **entire life-cycle**

Life-Cycle model



- Life-cycle model differs slightly depending on the context of institution and resource type
- This is a generic library model

Why a life-cycle approach?

- Digital materials are susceptible to change for different reasons throughout their life-cycle
- Stages impact on subsequent stages
- Enables continuity and provenance despite technological and organisational contextual change
- Maximises investments and potential
- Vital for reliable re-use

Planning

- Know what you've got
 - Inventory of resources
 - Risk assessment
- Identify where the responsibilities lie
- Get appropriate people involved
- Get institutional buy-in
- Develop policy & strategy
 - Different strategies for different resources
 - Different resources have different requirements

Strategy (1)

- A written policy and strategy to support activities and help secure resources
 - Take a life-cycle approach to support curation and preservation planning
 - If creating resources, provide good practice guidance for sustainability (eg when digitising or accepting digitised resources)
 - Assess collection/selection criteria – are they still valid? Do they need expanding? Identify possible resources
 - Digital resources can complement & enhance physical ones

Strategy (2)

- Identify legal restraints in collection/management/access
- Consider if value can be added to resources during acquisition?
- Store objects in a secure environment
- Plan for preservation activities to maintain access to authentic resources over time and avoid incurring extra costs
- Determine access and user requirements
- Re-visit and assess strategy regularly!

Practical Tips & Helpful Initiatives

Develop appropriate infrastructure

- Make solid business case
- Implement Institutional Repository
 - In house
 - External/outsourced
- Use standards
- Get trained up!
- Consider preservation services
- Use available tools to help you
- Audit (& Certification?)



RSP Summer School



Preserv Preservation Eprint Services

OpenDOAR Policies Tool

DRAMBORA

Avoid format obsolescence

- Store data in standard formats where possible
- Describe context with metadata
- Technology Watch
- Test strategy before implementing
- Maintain digital version – don't simply 'print to paper'

Migration: conversion of data into current or more widely accessible formats

Emulation: use of modern HW/SW to recreate old computing environment and run old/obsolete files

Protect physical storage media

- Environmental controls
- User controls!
- Regular refreshment
- Backups
- High quality materials
- Good documentation
- Quality control checks



‘Leased’ resources

- Find out publishers’ archiving policies
 - How long do they promise to retain old versions?
 - How long will they provide access to old content?
 - What is their preservation strategy?
- What if you cancel your subscription?
 - Do you still get access to ‘old’ content?
 - Can you retain copies of ‘old’ content?
- What’s the difference between the online version and the CD-ROM?
- Will corrupt CD’s be replaced?

Digital Curation Centre

- Established to **help solve the extensive challenges** of digital preservation and curation, and to **provide research, advice and support services** to UK institutions
- Funded by JISC & eScience Core programme
- 4 main areas of work:
 - Community Development
 - User Services
 - Research
 - Tools & Development
- Centre of excellence in digital preservation and curation

UK WAC



- UK Web Archiving Consortium (6 members)
 - British Library, National Library of Scotland, National Library of Wales, The National Archives, Wellcome Library, JISC
- Collects Web content selectively
 - Uses modified PANDAS collection/harvesting software developed by the National Library of Australia
 - Underlying harvesting program is currently HTTrack
 - Permission is sought from site owners in advance
 - Persistent Identifier URLs
 - Single partner assumes responsibility for each site
 - Central repository of metadata
 - The collections are publicly accessible
- Website: <http://www.webarchive.org.uk/>

Internet Archive



- Non-profit organisation, based in U.S.
- Wants to offer permanent access to digital online materials of all types
- Founded in 1996, has been collecting since then ... much content donated by Alexa Internet
- Collects sites by crawling and harvesting web sites
 - Sites can 'opt out' by way of robots.txt file on the web server
- Most content is freely available to the public, e.g. through the Wayback Machine
- Interface issues: only the URL indicates that the page is archived
- Website: <http://www.archive.org/>

IIPC

netpreserve.org
international internet preservation consortium

- International Internet Preservation Consortium
 - Builds co-operation between the Internet Archive and national and research libraries
 - Co-ordinated by the Bibliothèque nationale de France
 - The British Library is the only current UK member, other national library partners include the Library of Congress, the Library and Archives Canada and the national libraries of Australia, Denmark, Finland, Iceland, Italy, Norway and Sweden
 - Reflects those with current experience of Web archiving
 - Developing IIPC toolkit with standards and tools for supporting acquisition, collection management, storage & maintenance, and access & finding aids
- Website: <http://netpreserve.org/>

LOCKSS (1)

- Lots of Copies Keeps Stuff Safe (LOCKSS)
- An 'easy and inexpensive way to collect, store, preserve, and and provide access to their own, local copy of authorised content they purchase'
- E-Journal collection and preservation system
- Open Source Software
- Runs on standard desktop hardware
- Requires very little technical administration



LOCKSS (2)

- Trial and pilot projects underway
 - DCC support available through helpdesk and dedicated Advisory post
- Current trial suitable only for certain titles (due to licensing arrangements with publishers)
- Private networks *can* be developed:
 - Requires technical development
 - Minimum of six machines necessary to achieve desired redundancy
 - Suitable for, eg, online course material



Thank you.

Questions?

Maureen Pennock

m.pennock@ukoln.ac.uk

(Join the DCC Associates Network at
<http://www.dcc.ac.uk>)