Preservation Planning at the British Library

Michael Day
Digital Preservation Manager
The British Library
Preservation Planning at the British Library

Source: https://www.flickr.com/photos/britishlibrary/11227122385
Outline

• Contexts: the British Library and its digital collections
• Digital preservation challenges
• Understanding “preservation planning”
• Approaches to digital preservation at the British Library
• Developing preservation planning capacity at the British Library
The British Library and its collections
The British Library in history

Source: https://www.flickr.com/photos/britishlibrary/11117856633/
The British Library today
Magna Carta

Discover the history and legacy of one of the world's most celebrated documents

Examine collection items such as the British Library's 1215 copy of Magna Carta together with a translation in modern day English. Read an introduction to Magna Carta, learn more about the context of its creation, the people involved or explore a timeline of its legacy.

Magna Carta
Law • Liberty • Legacy

Exhibition open from 13 March – 1 September 2015
BOOK NOW

Articles
Magna Carta, 1215

Manuscript

John, by the grace of God King of England, Lord of Ireland, Duke of Normandy and Aquitaine and Count of Anjou, to his archbishops, bishops, abbots, earls, barons,
Current strategy: Living Knowledge

The British Library’s purposes

1. Custodianship
   We build, curate and preserve the UK’s national collection of published, written and digital content.

2. Research
   We support and stimulate research of all kinds.

3. Business
   We help businesses to innovate and grow.

4. Culture
   We engage everyone with memorable cultural experiences.

5. Learning
   We inspire young people and learners of all ages.

6. International
   We work with partners around the world to advance knowledge and mutual understanding.
Current strategy: Living Knowledge

Custodianship
We build, curate and preserve the UK’s national collection of published, written and digital content

“… our first and core purpose, the one on which all the others depend.”
Digital collections at the British Library
Digital collections at the British Library

- Non-Print Legal Deposit
  - From 6th April 2013, legal deposit provisions cover material published digitally and online, “so that the Legal Deposit Libraries can provide a national archive of the UK's non-print published material, such as websites, blogs, e-journals and CD-ROMs.”
  - [http://www.bl.uk/catalogues/search/non-print_legal_deposit.html](http://www.bl.uk/catalogues/search/non-print_legal_deposit.html)
Digital collections at the British Library

• Digitisation programmes, e.g.:
  – British Newspaper Archive; DC Thomson Family History partnership
  – Google Books partnership
  – Early English Books Online (EEBO); ProQuest partnership
  – eTheses (EThOS)
  – Manuscripts: http://www.bl.uk/manuscripts/
  – Maps
  – Sound and moving images

• Save our Sounds:
  – Project information: http://www.bl.uk/projects/save-our-sounds
Digital collections at the British Library

- Over 40 million pages of digitised newspapers
- 500,000 pages of India Office records and Arabic science manuscripts
- Over 40,000 digitised sound recordings
- 250,000 digitised books
- > 1 billion webpages
- > 500,000 e-theses
- Digitised & born-digital maps
- eJournals
- Personal digital archives
- Digitised treasures (e.g., Beowulf, Lindisfarne Gospels, Magna Carta, Shakespeare First Folio)
- Digitised stamps, patents, pamphlets, musical scores...
- Over 500 TB of content, with much more on the way...
The challenge of long-term digital preservation
What is the problem?

“Digital resources last forever or five years, whichever comes first”
Media deterioration
Technological obsolescence
Integrity: bit rot

• Damaged JPEG: http://www.flickr.com/photos/tjvcarter/8620812799/
Authenticity: file rendering

Figure 28: Corel WordPerfect version 7 file rendered in Corel WordPerfect version 7

Figure 29: Corel WordPerfect version 7 file rendered in LibreOffice Writer 3.3.0
Digital preservation

“Digital preservation is the series of actions and interventions required to ensure continued and reliable access to authentic digital objects for as long as they are deemed to be of value.”

http://www.jisc.ac.uk/publications/briefingpapers/2006/pub_digipreservationbp.aspx
Digital preservation

• Is concerned with the *sustainability* of “content” for a given period of time (not always forever)

• Is about ensuring “*continued access*” to content

• Is a *managed series of activities*, including:
  – The identification and adoption of appropriate preservation strategies
  – The collection and management of appropriate metadata
  – The ongoing monitoring of technical contexts
  – The ongoing monitoring of the organisation (audit)

• Combines technical, organisational and legal challenges
Preservation planning
Preservation planning - definitions

• Northeast Document Conservation Center definition:
  – “Preservation planning is a process by which the general and specific needs for the care of collections are determined, priorities are established, and resources for implementation are identified.
  – Its main purpose is to define a course of action that will allow an institution to set its present and future preservation agendas.
  – In addition, it identifies the actions an institution will take and those it probably will never take so that resources can be allocated appropriately.”

https://www.nedcc.org/free-resources/preservation-leaflets/1.-planning-and-prioritizing/1.1-what-is-preservation-planning
Preservation planning - definitions

• Reference Model for an Open Archival Information System (OAIS), ISO 14721:2012
  “The OAIS functional entity which provides the services and functions for monitoring the environment of the OAIS and which provides recommendations and preservation plans to ensure that the information stored in the OAIS remains accessible to, and understandable by, and sufficiently usable by, the Designated Community over the Long Term, even if the original computing environment becomes obsolete.” (p. 1-14)
Preservation planning - definitions

• Covers a very wide range of digital preservation activities
  – Many different approaches; needs to be organisation-specific
  – Built upon requirements gathering

• Typically focuses on:
  – Collections – e.g. the evaluation of content, the identification, development and adoption of appropriate preservation strategies, recommendations on standards, risk analysis, quality measures, validation, etc.
  – Organisations – e.g., policies, resource allocation, risk analysis, maturity levels, etc.
  – The wider environment – e.g., monitoring changes in technology and user needs, etc.
Preservation planning - tools

• Content focused:
  – Plato Preservation Planning tool
  – Data Asset Framework (research data management)

• Organisation focused:
  – The NDSA Levels of Digital Preservation
  – Trusted Digital Repositories (TRAC, ISO 16363)
  – DRAMBORA (risk assessment framework)
  – CARDIO (benchmarking tool for research data management)

• Environment focused:
  – Scout (preservation watch)
Plato Preservation Planning tool

Objects
Technology
Usage criteria
Policies
Actions

Preservation planning environment

Define requirements
Evaluate alternatives
Analyse results
Recommendation
Build preservation plan

Knowledge base

Monitor
- requirements
- technology
- environment

Repository

Preservation plan

http://www.ifs.tuwien.ac.at/dp/plato/intro/
Digital preservation at the British Library
Storage Infrastructure

Storage nodes:
- BL St Pancras (STP)
- BL Boston Spa (BSP)
- National Library of Wales (NLW)
- National Library of Scotland (NLS)

Access gateways:
- Bodleian Libraries (Ox)
- Cambridge University Library (Ca)
- Trinity College Library, Dublin (Tr)
Who is involved?
“By 2020, end-to-end workflows are in place that deliver and preserve our digital collections in a trusted long term digital repository so that they may be accessed by future users.”
A three-tiered approach
Digital Preservation Principles I

1. We integrate curatorial assessments of our digital collection content into preservation decisions *so that* technical activities support curatorial requirements for the collections

2. We preserve metadata about our digital collections *so that* we may understand and preserve the collections over time

3. We preserve the provenance of our digital collection content *so that* we understand and can demonstrate its authenticity over time

4. We record any modifications to digital collection content (e.g. preservation action, normalisation) during the lifecycle *so that* we can understand and demonstrate its integrity over time

5. We consistently apply and document our application of metadata standards *so that* future generations can understand our collections

6. We maintain file-level integrity of our digital collections *so that* we can protect against loss and damage
Digital Preservation Principles II

7. We preserve original files in our long term repository, alongside any other required representations of the content, so that we maintain the original artefacts acquired or deposited into our care as a ground truth representation of the content for future, currently unknown, preservation and access scenarios.

8. We maintain Preservation Master copies of collection content in our long term repository so that the format-based risks of preservation over time are minimised.

9. We maintain and implement preservation plans for our digital collections so that preservation actions are reliable and based on a holistic understanding of the collections and their context.

10. We implement comprehensive end-to-end workflows so that we may consistently manage and preserve our digital collections across the entire lifecycle.

11. We regularly monitor our digital collection content for emergent preservation risks so that we may mitigate against them.

12. We integrate quality assurance checks into the lifecycle where appropriate so that the authenticity and integrity of the content is maintained.
Developing requirements

• Digital preservation is business as usual

• Principles for developing a preservation planning capacity:
  – Incremental (no need to do everything at once)
  – Practical (e.g. focused on institutional requirements)
  – Flexible
    • No one-size-fits-all solution (although needs to operate at scale)
    • Needs to cover a wide range of content types
    • Takes into account the whole life-cycle
    • Other stakeholders (e.g. legal deposit libraries)
  – Evidence-based (e.g., assessment-based activities)
Preservation planning at the British Library
Preservation planning and collection care

• Evidence-based collection care principles
  – PAS 198:2012 Specification for managing environmental conditions for cultural collections
  – Deborah Novotny (IFLA Conference 2012, Helsinki):
    • “It is not sufficient to say that the collection will be preserved in perpetuity. Rather, it is necessary to assess the significance of the collection, or each section of the collection, deciding what features of the collection should be preserved, for what reasons”
    • «Il ne suffit plus d’affirmer que la collection sera préservée à perpétuité. Par contre il est nécessaire d'évaluer la valeur patrimoniale de chaque collection, ou de chacune de ses sections, en décidant quels éléments de la collection doivent être préservés et pour quelles raisons.»
Collection profiling

• What content do we have and what is important about it?
  – Starting with existing Digital Asset Register

• Collection profiling:
  – Documenting key knowledge about the Library’s high-level digital collections
  – A mechanism for liaising with curators and collection specialists
  – Considering preservation requirements / preservation intent for these collections
Collection profiling

• What is Preservation Intent?
  – A summary of points, agreed by curators, identifying the main characteristics of collections that must be preserved.
  – *D-Lib Magazine* paper by Colin Webb, David Pearson, and Paul Koerbin of the National Library of Australia:
    [http://www.dlib.org/dlib/january13/webb/01webb.html](http://www.dlib.org/dlib/january13/webb/01webb.html)
Collection profiling

• The profile framework (has evolved over time):
  – Summary
  – What is this format, and what is it within the Library?
  – Acquisition
  – Preservation Intent
  – Issues
  – Summary table of sub-collections

• Collection profiling is a first step in an ongoing process of capturing preservation requirements
File format assessments

• Sustainability assessments of formats
  – Interacts with ongoing discussion on “preservation masters”
  – Less concerned with theoretical “obsolescence,” but aware of institutional factors in maintaining access to content
  – Principles:
    • Based, where possible, on real-world evidence (documented examples)
    • Clarity on purpose and intended audience
    • Not based on numerical scoring measures
File format assessments

- Main criteria
  - Development Status
  - Adoption and Usage
  - Software Support
  - Documentation and Guidance
  - Complexity
  - Embedded or attached Content
  - External Dependencies
  - Legal Issues
  - Technical Protection Mechanisms
  - Other …
Developing preservation planning capacity

• Other tasks:
  – Tool assessment
  – Workflow assessment (considering content lifecycle)
  – Policies
  – Training
Planning for both scale and complexity

- eMSS
- Digital Mapping
- Web archives
- eJournals
- eBooks
- eTheses
- Digitised Sound
- Digitised MSS
- Digitised Newspapers
The Library’s external collaborations:

Research projects:
Thank you

Michael Day
Digital Preservation Manager
The British Library
96 Euston Road
London NW1 2DB
United Kingdom
michael.day@bl.uk

https://www.flickr.com/photos/britishlibrary/11004654615/
Links – The British Library

British Library website: http://www.bl.uk


British Library, Non-Print Legal Deposit: http://www.bl.uk/catalogues/search/non-print_legal_deposit.html

British Library, Save our Sounds project: http://www.bl.uk/projects/save-our-sounds

British Library, Digital Preservation Strategy, 2013-2016:


Links – Digital preservation (1)


Reference Model for an Open Archival Information System (OAIS), CCSDS 650.0-M-2: http://public.ccsds.org/publications/archive/650x0m2.pdf

Plato Preservation Planning tool: http://www.ifs.tuwien.ac.at/dp/plato/intro/

Data Asset Framework: http://www.data-audit.eu/

Links – Digital preservation (2)

TRAC, Trustworthy Repositories Audit & Certification: Criteria and Checklist: 
http://www.crl.edu/sites/default/files/d6/attachments/pages/trac_0.pdf

Audit and Certification of Trustworthy Digital Repositories, CCSDS 652.0-M-1: 
http://public.ccsds.org/publications/archive/652x0m1.pdf

ISO 16363: ISO 16363:2012, Space data and information transfer systems -- Audit and certification of trustworthy digital repositories

DRAMBORA: http://www.repositoryaudit.eu/

Digital Curation Centre, CARDIO: http://cardio.dcc.ac.uk/

Scout preservation watch: http://openpreservation.org/knowledge/blogs/2013/12/16/scout-preservation-watch-system/

British Library MS Cotton Augustus II.106