

# **Digital and Content Strategy Developments at the British Library**

**John Tuck  
Head of British Collections**

**M25 Consortium of Academic Libraries Business Meeting  
January 23<sup>rd</sup> 2007**

# Digital World

“In the new digital world we see a picture of ever more rapid innovation, much of which is happening outside libraries and driven from the commercial sector, a picture of confusion and contradiction in the range of business models that are emerging and being experimented with; and new demands from discerning and empowered users”.

Lynne Brindley, Chief Executive, British Library

# BL Strategy and Digital Developments

Redefining the Library, British Library strategy, 2005-09:  
Strategic Priority 2: Build the digital environment

- Legal Deposit Libraries Act, 2003: extension of legal deposit to non-print
- Web Archiving
- Digitisation: Strategy
- Digitisation: Projects
- Content Development: Print-Digital Transition

# Legal Deposit: Extension to Non-Print

- Enabling legislation now moving to secondary legislation through the work of the Legal Deposit Advisory Panel (LDAP)
- Work carried out on mapping the universe of electronic publications potentially eligible for legal deposit
- Sub-groups of LDAP focussing on specific areas: offline; e-journals; web-sites

# Legal Deposit and Websites: Progress

Proposal being considered by LDAP as follows:

- Regulation to be put in place to enable legal deposit from the public domain web space of the UK through a combination of domain-level harvesting and selective web archiving
- Retention of such content by the Legal Deposit Libraries to be enabled
- Access to the content within the Legal Deposit Libraries infrastructure to be enabled and not hindered by legal barriers such as FOI
- Technical mechanisms to be put in place to make deposit easier and more effective

Aims:

- Regulation to come into force as soon as is practicable
- Impact assessment and consultation to be wide enough to include all relevant publisher groupings

Realistic timescale:

- Not likely to be implemented before 2008

# UK Web Archiving Consortium: UKWAC

- UKWAC comprises six institutions: British Library, National Library of Scotland, National Library of Wales, The National Archives, JISC, and the Wellcome Trust.
- Initial two-year project to develop and evaluate a collaborative infrastructure for web archiving in the UK
- Has achieved its objectives and has agreed extension until at least September 2007
- UKWAC operates on website owner rights clearance basis: less than 25% successful return rate on permissions; although few outright refusals
- Live archive of web sites, fully accessible, free of charge is in place <http://www.webarchive.org.uk>. Over 1,800 sites available

# UKWAC: Collecting

Significant thematic collections strand: seven examples:

- General election UK 2005 (93 sites)
- Personal experiences of illness (27 sites)
- Tsunami disaster 2004 (23 sites)
- Avian and pandemic influenza (11 sites)
- Digital lives (24 sites)
- Terrorist attacks, London, July 7th 2005 (49 sites)
- Women's issues (117 sites)

More underway, e.g. Olympics 2012

# International Developments with International Internet Preservation Consortium (IIPC): Automated Smart Crawler

## Automated Smart Crawler

- Joint project involving British Library (project lead), Library of Congress, Bibliotheque nationale de France and Internet Archive
- Scope includes developing Heritrix crawler so that it has ability to:
  - recognize when resources have not changed from last time they were gathered
  - Prioritise the order in which resources are visited and gathered
  - Recognize when resources are changing more frequently and to visit them more frequently
- Ensure that new capabilities will scale to crawls of at least 100 million resources
- Ensure that web archiving community can take advantage of these new capabilities

# Digitisation at the British Library: Background

- Digitising for more than ten years
- Originally three main categories:
  - Heritage items, e.g. Beowulf, Gutenberg Bible
  - Larger scale projects where aim is to create a more comprehensive resource that can be viewed as a single collection, e.g. Collect Britain
  - Commercially funded projects that digitise parts of the collection and then sell or license the digital collection, e.g. Early English Books Online

# Current Strategy

- Focus on building large digital collections
- Focus on large collections of digitised newspapers
- Focus on large collections of publicly accessible audio material
- Development of software applications for e-learning, etc.
- Partnerships with other organisations to create larger virtual collections
- Fundraising for small digitisation projects to broaden access to the collections

# Significant Projects: British Newspapers 1800-1900 Digitization Project

## Project Objectives

- To digitise up to two million pages of British national, regional and local newspapers from microfilm
- To offer access to that collection via a searching and browsing interface on the Web
- Funded by Joint Information Systems Committee (JISC)
- Further bid for additional funding submitted in May 2006
- For more information see <http://www.bl.uk/collections/britishnewspapers1800to1900.html>

# Highlights of British Newspapers Project

The project will offer users in Higher Education (HE) and Further Education (FE):

- Free access to a virtual library of nationally, regionally and locally important digitised British newspapers from 1800 to 1900
- Free access to a broad range of valuable learning materials, many of them unique, and hitherto available only in London-based reading rooms, in hard copy or in microform
- The ability to search across the different newspaper titles to draw together materials relating to a wide range of research and learning topics

# Significant Projects: Archival Sound Recordings Digitization

- Project also funded by Joint Information Systems Committee
- Aim was to digitize 4,000 hours of content from the collections of the Sound Archive. Launched in September 2006. See <http://www.bl.uk/sounds>
- Bid for further funding submitted in May 2006

## Content includes:

- Beethoven String Quartets; 80 years of performance
- 400 popular music tracks; a teaching set reflecting trends and developments in presentation, recording techniques, production, marketing, lyric content, instrumentation etc

# Archival Sound Recordings: Content

- African field recordings
- Sony Radio Awards
- African Writers Club: broadcasts of literature reviews and commentary, politics and music
- Oral history of jazz in Britain
- Visual arts interviews and Architects and designers
- Records & Players:
  - contextual multimedia information about recording techniques
  - recording & playback developments
  - restoration and audio history

# Significant projects: Codex Sinaiticus

- Fourth-century Codex Sinaiticus is one of the two earliest Christian Bibles
- The Codex is the remains of a huge handwritten book that contained all the Christian scriptures of the old and new Testaments, together with two late first-century Christian texts, the Shepherd of Hermas and the Epistle of Barnabas
- The Codex is dispersed between four institutions:
  - St Catherine's Monastery
  - The British Library
  - Leipzig University Library
  - The National Library of Russia, St Petersburg

# International Codex Sinaiticus Project

A major international partnership has been agreed and was formally signed at the British Library on 9<sup>th</sup> March 2005.

## **Aims of the project include:**

- Full conservation, codicological and palaeographical examination of the Codex
- Conservation work to stabilise the Codex for digitization and preservation for the future
- Define and undertake scholarly work on the Codex and to deliver a range of electronic and printed products, including website
- To undertake research on the history of the Codex

# International Codex Sinaiticus Project

- Project overseen by international project board
- External funding sought for the project
- Successful funding awards and grants from:
  - Arts and Humanities Research Council
  - Deutsche Forschungsgemeinschaft
  - Niarchos and Leventis Foundations

# Out of Copyright Books: Partnership with Microsoft

- Microsoft and the British Library: a strategic partnership to digitise out-of-copyright books
- Selection Scope: British printed books; nineteenth century, avoiding material included in EEBO at ECCO
- Search results for this content are being delivered through MSN Live Book Search

# British Library's Content Strategy

- Consultation exercise took place, April to July 2006
- Focussed on c. £15m per annum acquisitions budget
- Summary of responses has been posted on the web (see <http://www.bl.uk/contentstrategy>)
- Broad consensus on many points, including recognition of the print-digital transition and transition from a collecting to a collecting and connecting (i.e. content) strategy

# Content Strategy: Managing the Print-Digital Transition

- British Library presented two categories of material for which a decision has been taken to change the collecting policy and shift to the purchase of digital copies only (wherever possible): purchased UK journals that duplicate print copies held under legal deposit; and purchased overseas journals
- Issues raised: ensuring long-term access and preservation; cost factors; implications for document supply
- Content Development Action Plan