DIGITAL CONTENT AND LIFE-CYCLE COLLECTION MANAGEMENT: A BRITISH LIBRARY PERSPECTIVE

Dr Clive Field

Director of Scholarship and Collections, The British Library

March 2003
OVERVIEW

• Profile of British Library and its collections
• Collection development policy
• Digital collection development
• Digital object management
• Life-cycle collection management
• Conclusions - critical success factors
HISTORY OF THE BRITISH LIBRARY

• Created by British Library Act 1972 - commenced 1973

• Merger of British Museum Library (1753), National Reference Library of Science and Invention (1855), National Central Library (1916), and National Lending Library for Science and Technology (1961)

• Subsequent incorporation of British National Bibliography in 1974, India Office Library and Records in 1982, and British Institute of Recorded Sound in 1983

• Flagship building at St Pancras - largest public building project in Great Britain in 20th century - opened in 1997
• 2.6 million items per annum consulted in reading rooms
• 2.9 million remote supply items per annum
• 100 millionth remote item in 2001
• 432,000 reading room visits in 2001/2
• 57,000 new reader passes issued in 2001/2
• 366,000 exhibition visitors in 2001/2
• 2,400 staff
• £83 million Grant in Aid and £28 million trading income
• Annual report - http://www.bl.uk/about/annual/latest.html
THE COLLECTIONS

• 150 million collection items

• 2.5 million items added each year

• 649 km shelf capacity and 92% full at end of 2001/2

• 11 km of material net added each year (new acquisitions less deaccessioning of duplicates)

• 5 km of legal deposit items added each year (526,000 items in 2001/2)

• 7 km of purchased items added each year (£14 million per annum)

• 187,000 books and 112,000 serial titles added in 2001/2
COLLECTION DEVELOPMENT POLICY (1)

• Patchwork of many smaller collections formed by organisations and individuals over several hundred years - thus not always easy to determine retrospective policy

• Comprehensive collection of UK national published archive in most formats, underpinned by legal deposit legislation for current material and by retrospective purchasing to fill gaps

• Universal, if no longer encyclopaedic, collection of overseas published material, underpinned by maintenance of acquisitions fund well in excess of purchasing power of any other UK institution, and enabling British Library to be gateway, acting in concert with others as appropriate, to “the world’s knowledge”
COLLECTION DEVELOPMENT POLICY (2)

• Extensive collection of unpublished materials, preponderantly from UK but also from overseas as appropriate, building upon existing strengths and taking account of collecting priorities of other libraries and archives

• Collection which is format-blind and hybrid, in that selection decisions are governed by collection development priorities for each area, but within overarching framework of life-cycle collection management and costing

• Where publication available in parallel print and electronic formats, paper is still preferred as the preservation format, although a digital subscription or licence may be taken out to support access
COLLECTION DEVELOPMENT POLICY (3)

• During course of 2003 series of policy statements, for publication, being produced to clarify British Library’s approaches to collection development

• These will cover, for example, UK national published archive, criteria for purchase of items in single copies, criteria for purchase of items in duplicate, and deaccessioning
DIGITAL VOLUNTARY AND LEGAL DEPOSIT MATERIAL (1)

• Voluntary scheme for deposit of commercial audio publications has operated for many years between British Phonographic Industry and British Library

• This is extremely successful and, since advent of CDs, has resulted in intake of large amounts of born-digital sound content

• Voluntary scheme for deposit of offline and microform publications, involving all six legal deposit libraries, instituted in January 2000, with Government blessing and following Kenny Report on extension of legal deposit

• Scheme administered by Joint Committee on Voluntary Deposit (JCVD), and supported by published Code of Practice
DIGITAL VOLUNTARY AND LEGAL DEPOSIT MATERIAL (2)

• Scheme has been relatively successful, attracting participation of over 100 publishers, who are increasingly depositing online material, especially PDFs

• Estimated that 90,000 digital items will have been deposited with British Library under scheme by end 2004

• Both library and publisher sides of JCVD, and Government, agree that only statutory scheme will guarantee integrity of national published archive, and that time is ripe for legislation

• Regulatory impact assessment prepared by Government, quantifying benefits and costs for libraries, publishers, and nation over ten years - informed by research for JCVD from Electronic Publishing Services
• Private Members’ Bill to extend legal deposit to non-print publications currently before Parliament; it is generic, with application to specific information formats coming through subsequent Orders in Council

• For commercial digital publications British Library will aspire, over time, to build up to somewhere near comprehensiveness of coverage achieved for print legal deposit, with early priority given to e-only publications

• For online materials likely implementation approach will be single copy deposit and secure networking across all legal deposit libraries - proof of concept network already built and demonstrated to publishers
UNITED KINGDOM WEBSITES (1)

• British Library’s Domain.UK pilot project in 2001/2 covered representative sample of 100 UK sites (from universe of 2.7 million) across Dewey Decimal Classification

• These were ‘whacked’, with rightsholders’ permission, every three weeks using Bluesquirrel software in order to gain first-hand experience of technology, selection and management issues

• British Library bid to Government (initially Comprehensive Spending Review 2002, now Invest to Save) to commence UK Web harvesting programme

• Proposal based upon annual snapshot of UK domain and selective harvesting of up to 10,000 research-interest sites
• Outcome of bid, for £600k per annum (with separate bid for storage), still awaited, but British Library will commit £100k per annum from its resources from 2003/4

• British Library will join international consortium of national libraries (including Library of Congress, Bibliothèque nationale, and others) to agree distributed global approach to Web archiving

• Consortium likely to be working with Internet Archive to develop next-generation Web crawler and to acquire retrospective content from 1996

• British Library refining collection development policy for UK Websites, taking account of potential for collaboration with other UK partners (Joint Information Systems Committee, Public Record Office, Wellcome Trust)
PURCHASED DIGITAL MATERIAL

• Annual expenditure of around £1.25 million on UK and overseas commercial digital materials

• Purchase and licensing for reading room use (standalone or networked) of e-only publications consistent with collection development policy - see http://www.bl.uk/collections/wider/eresources.html

• Licensing of parallel e-formats of high-demand publications already held in print - for service reasons, for reading room and, especially, remote users

• 4,000 journals already licensed for inclusion in e-access store to facilitate document supply

• British Library has pioneered secure electronic delivery with Elsevier and Adobe, on a copyright fee paid basis
DIGITISATION (1): KEY SELECTION CRITERIA

- Likelihood of widening access to collections
- Likelihood of reducing handling of fragile originals
- Evidence of actual or potential demand
- Likelihood of addressing dual (scholarly and public) purpose
- Development of critical mass of related material
- Value-added features, not just images (e.g., searchable text)
- Extent to which dispersed materials will be virtually reunited
DIGITISATION (2): KEY SELECTION CRITERIA

• Extent to which cultural restitution issues addressed

• Extent to which service and property strategies supported

• Maintenance of appropriate balance between public good (freely available) and commercial (chargeable) resources

• Minimisation of call upon Government Grant in Aid, to avoid siphoning resources from traditional core

• Avoidance of duplication with work of other digitisation programmes

• Minimisation or avoidance of IPR issues

• Likelihood of sustainability
DIGITISATION (3): SOME CURRENT PROJECTS

• International Dunhuang Project - succession of grants from 1993, with 31,000 images - http://idp.bl.uk

• In Place - grant from New Opportunities Fund (Lottery), with 100,000 objects (text, sound, image)

• English books before 1801 - commercial projects with Proquest and Gale, based upon pre-existing microfilm editions

• Burney newspapers - grant from National Science Foundation

• 19th-century British newspapers - grant from Joint Information Systems Committee
DIGITISATION (4): SOME CURRENT PROJECTS

• Sound Archive - grant from Joint Information Systems Committee and internally-funded programme to transfer vulnerable analogue materials to CD-Rs

• Digital scholarship - wide range of mainly externally-funded digital research projects (including digitisation), several funded by Arts and Humanities Research Board

• Treasures Gallery - wide range of mainly externally-funded digitisation projects of heritage items, eg partnership with Keio University to digitise Gutenberg Bible and both editions of Caxton’s printing of *Canterbury Tales*, and with Octavo to digitise Shakespeare Quartos

• Turning the Pages - innovative British Library-developed application, now being commercialised with Armadillo
• Long-term storage and preservation of, and access to, majority of British Library digital assets is current goal

• All voluntary and legal deposit UK material (including Websites) to be secured

• Digitised masters created from British Library collections to be secured (but subject to some debate)

• Some minority of overseas material to be secured, dependent upon preservation strategies of other libraries

• Overall aim of DOM programme to preserve, with backup copies, and provide access to (within appropriate permissions), original bit-stream of digital objects and associated metadata
• Original physical media and packaging will be saved as artefacts but generally not used as part of digital preservation strategy

• Minimum set of metadata consistent with long-term preservation and access will be constructed

• Special access versions will be created selectively, to allow items to be viewed via current desktop technology

• Process covers three stages: collection (selection, acquisition, accession, cataloguing), preservation (storage, preservation) and access (discovery, searching, retrieval, delivery)
DIGITAL OBJECT MANAGEMENT (3)

• Interim phase of programme, to 2004/5, involves
  a) maintenance and scaling up of existing hybrid storage and access solutions for commercially-produced sources
  b) archiving to DVD-R of digitised masters from British Library collections
  c) creation of internal write once, read only preservation store with limited access for UK material received under voluntary deposit, utilising Ex Libris DigiTool for organisation and management

• Second phase of programme, from 2004/5, involves creation of more comprehensive and integrated solution, based upon collaborative approach - discussions under way with range of partners in UK, US and rest of world
LIFE-CYCLE COLLECTION MANAGEMENT (1)

• Aims to produce report, recommendations and implementation strategy for policy and economic framework for life-cycle collection management of all information formats

• Task 1 - Definition: define various phases in life-cycle, revalidate or revise earlier costing formula for print monographs and periodicals, develop formulae for e-life-cycle (separate for voluntary and legal deposit, digitised masters, and purchased electronic resources)

• Task 2 - Internal data gathering: analysis of how much being spent on each phase, identification of gaps in management information, identification of backlogs as indicators of mismatch of resources
LIFE-CYCLE COLLECTION MANAGEMENT (2)

• Task 3 - External review: review literature and national and international practice for library, archive and heritage sector, and for other sectors, and identify collaborative opportunities and benchmarks.

• Task 4 - Internal linking: produce matrix of managerial, service and economic inter-dependencies between various stages of life-cycle, and ensure that outputs of project elide with corporate strategy and business plans.

• Outcomes of project may lead to recalibration of resourcing between and within traditional and e-life-cycles (for instance, acquisitions vs. metadata) and to reprioritisation of what is ultimately selected for long-term digital preservation.
CONCLUSIONS - EIGHT CRITICAL SUCCESS FACTORS

• Clear long-term collection development strategy
• Clear long-term preservation and access strategy
• Clear and sustainable business models, balancing public good and commercial
• Clear and sustainable service models
• Standards compliance (interoperability and metadata)
• National and international partnerships with libraries, rightsholders and others
• Supportive legislative environment
• Supportive Government (for national libraries)