Collecting, selecting and legal deposit

John Tuck
Head of British Collections

DPC Forum on Web Archiving
British Library, 12th June 2006

Legal Deposit:
Legal Deposit Framework in the UK

- Origins of Legal Deposit: France, 1537; Thomas Bodley, 1611

- 1911 Copyright Act giving legal deposit entitlement to six libraries:

- British Museum; Cambridge University Library; Bodleian Library; National Library of Wales; Faculty of Advocates Library, Scotland; Trinity College, Dublin

- 1999 Voluntary Code of Practice/Joint Committee on Voluntary Deposit

- Legal Deposit Libraries Act 2003

Legal Deposit:
Legal Deposit Libraries Act 2003 (1)

- Enabling legislation

- Did not change legislation for printed publications

- Did not change entitlement of the existing libraries

- Did extend to non-print

Legal Deposit:
Legal Deposit Libraries Act 2003 (2)

Secretary of State given power to make Regulations to cover:

- offline/online formats including associated computer programs to provide access to the works

- rights of use, especially as regards reader access

- provision for libraries to harvest material from the internet

- in summary, the framework within which the collection, capture, preservation and access to non-print materials can be set up

Legal Deposit:
Legal Deposit Libraries Act 2003 (3)

- Legal Deposit Advisory Panel (LDAP)

- Established by the Department of Culture, Media and Sport in 2005 (DCMS)

- Purpose is to advise the Secretary of State on the timing and content of Regulations relating to legal deposit

Legal Deposit:
Legal Deposit Libraries Act 2003 (4)

- More specifically aims of LDAP to:

  - LDAP comprises a Chair and 14 representatives from the legal deposit libraries, from publishers, together with a number of independent members

  - Initial focus: offline publications; e-journals; web sites; definition of UK territoriality; and overall workplan and procedures

  - advise the Secretary of State on the content of Regulations for the deposit of particular classes of non-print media, making recommendations for future secondary legislation as appropriate

  - monitor and oversee the operation and implementation of the Legal Deposit Libraries Act 2003

  - execute all these responsibilities in accordance with the letter and the spirit of the 2003 Act

  - advise the DCMS as requested on the operation of the 2003 Act

Legal Deposit:
Legal Deposit Libraries Act 2003 (5)

- Joint Committee on Legal Deposit (JCLD): forum for publishers and legal deposit libraries to discuss, collaborate on and resolve legal and voluntary deposit issues, both in advance of and subsequent to Regulations coming into force; to share information with, and offer advice to, the Legal Deposit Advisory Panel and the government on legal deposit issues of perceived statutory significance

- JCLD has operated working groups:

  - Review of voluntary scheme for offline publications

  - Territoriality: definition of UK

  - E-Journals and e-journals pilot

Legal Deposit:
Legal Deposit Library Act 2003 (6)

Legal Deposit Libraries Committee

- Comprises the Librarians of the UK and Ireland legal deposit libraries

- Also has four sub-groups: Collection Development; Metadata; Preservation; Technical Infrastructure

- Sub-Groups to be transformed into a Legal Deposit Libraries Committee Implementation Group: representation from all the legal deposit libraries and goal of taking forward the tasks to achieve implementation

Territoriality: Defining the UK

- Objective of legal deposit in any country is to acquire, preserve and give access for the long term to the cultural, intellectual and national heritage of that country, produced and distributed in different formats and editions

- How does this translate into an e-environment?

- National boundaries less relevant than in the paper-based environment?

Collection Development Policy:
British Library

Web Archiving High Level Collection Development Policy

Given the huge scale and dynamic nature of the web (estimated in 2004 at approx 5 million UK-based web sites) the British Library does not consider it practicable nor affordable to aim at truly comprehensive coverage of the UK web presence. The Library's strategy is based on:

a) taking a complete snapshot of the entire UK web presence at regular intervals (possibly annually or twice a year)

b) achieving a more intensive and selective harvesting of a limited number and well-defined range of sites, building up over time to perhaps 10,000. These would be sites judged to be of research value now and in the future, reflecting the national and cultural heritage, and including a number of sites which are exemplars of web innovation. Also included is an events-based, thematic collection strand.

Web Archiving Collection Development Progress

- Through its Curator, Web Archiving, the British Library has defined a more detailed development policy statement for UK web sites (See www.bl.uk/collections/britirish/modbritcdpwebsites.pdf)

- See also *The British Library's Content Strategy: Meeting the Knowledge Needs of the Nation*, Appendix 2, Websites; 'The Library will increase its collecting of UK websites in line with legal deposit legislation and technological feasibility. It will collect UK websites extensively through regular whole domain snapshots and selectively (in more depth and with greater frequency) for sites of UK research value.'

- Consultation document; responses to be received by July 21 2006. See www.bl.uk/contentstrategy

Web Archiving Collection Development: Selection

- Framework of curators within the British Library to assist the Curator, web archiving

- Longer-term aim is to consider web-sites as just another format to collect within an overall collection development policy and to embed within overall selection guidelines for content experts

- Work also carried out with partners, e.g. within the UKWAC consortium

Collection Development:
Policies of UKWAC Partners (1)

- **Joint Information Systems Committee**: to collect sites created as a result of funding awards made by JISC and sites relevant to the UK HE sector.

- **The National Archives**: focus on six main clusters of government departments: defence and foreign policy; administration of justice and internal security; management of national resources; provision of services not provided by the market (health, education, culture); regulation and coordination of market provided services; machinery of government and delegated administrations.

Collection Development:
Policies of UKWAC Partners (2)

- **National Library of Scotland**: to collect websites that reflect the knowledge, culture and history of Scotland and the Scots.

- **National Library of Wales**: to create a collection that provides a reflection of life in contemporary Wales whilst also serving the current and future research needs of Wales.

- **The Wellcome Library**: to collect health and medicine-related sites from the UK domain.

Collection Development:
UKWAC Achievements

- We have www.webarchive.org.uk

- 1,388 sites currently available for access; c. 4,000 instances; free to everyone

- UKWAC Collection development sub-group has:
  - put in place a framework of selection policies to ensure avoidance of duplication and overlap;
  - sought to develop best practice, e.g. from events-based archiving, i.e. collecting general election; tsunami; London bombing sites;
  - addressed issues around frequency and depth of capture of sites
  - has modified collaboratively the subject headings linked to the PANDAS software
  - has addressed unforseen issues, e.g. kitemarking (does archiving by UKWAC imply some form of quality recognition; concerns of website owners about archival sites being mistaken for current sites; etc)

Collection Development:
Barriers to Progress

- Here are five challenges:
    - What is UK webspace?
    - Different scale of collecting; hard to determine a representation range of selected items not of a universe of millions
    - Archiving technology has fallen behind website technology – not always possible to collect what is in range for technical reasons
    - New collaborations required – cannot do this alone
    - Permissions: sites only collected once explicit permission has been given by website owner. Low response rate for British Library, i.e. 25%; therefore labour intensive and risk of unbalanced collection

Collection Development:
Overcoming Barriers

What would facilitate collection development and selection?:

- remove the lengthy, time-intensive permissions process, which would:
  - allow domain-extensive archiving
  - allow rapid and deeper response to events
  - free more time for curation of the growing collections, including creation of subject-related collections and development of targeted resource-discovery, promotion of the collections, including development of liaison with the research community

Collection Development:
Overcoming Barriers (2)

- In a further phase, use smart-crawler technology to be programmed with elements of the collection development policy (and more detailed subject guidelines), to select subject-related segments (but large segments) of the domain, as set out by curator input

- offer quickly a demonstrably large archive available via the web itself