

Scholarly communications and digital strategy

Lynne Brindley

CEO, The British Library

April 2005

Agenda

- Key trends in scholarly communications
- UK developments
- BL initiatives
 - E-infrastructure development
 - E-legal deposit
 - Web archiving
 - Digital archiving services
 - Digital access to research articles

This is what I will cover in this talk

Scholarly communication – key trends

- **Acceleration in pace of change from print to digital**
 - Emergence of powerful, global networks
 - Supported by standards and interoperability
- **Improvements in online navigation**
 - Google Scholar
 - Scopus, Web of Knowledge
- **Business models and the ‘Open movement’**
 - Golden, green and grey routes to access
 - UK Government / Research Councils position on the issue of open access

Digital technology transforming the nature of research & enabling new kinds of digital publishing:

- greater volumes of data - size of the internet estimated by Berkley research at 500,000+ terrabytes; estimates suggest that the amount of scientific / technical information doubles every 5-7 years (cf. Rick Luce Trieste presentation); over 60,000 non-print items were published in the UK last year – most in STM
- distributed research networks & increasing number of researchers who want / expect to be able to access to the information they need via the desktop

•Google - leading pioneer in online navigation - 200 million searchers per day and 61 million unique users per month.

•Transforming the landscape through a number of new ventures – e.g. Google Print, its library digitisation programme and Google Scholar - launched late 2004

•Publisher recognition of the importance of navigation services for scholars – Elsevier has launched its own generic resource discovery tools – Scirus and Scopus

New technology enabling new forms of scholarly publication / communication & new business models

‘Golden Road’ - Author pays (e.g. BioMed Central – currently publishes 100 titles)

- Public Library of Science seeking publication fees of \$1,500 per article, launching series of high-prestige titles, supported by Moore foundation funding

‘Green road’ - Institutional Repositories - Authors self-archive their articles.

•Central support in some countries for pursuing this new model (UK – FAIR programme; Netherlands – DARE; Australia – ARROW). During the past two years, several OA journals have received subsidies from JISC and recently up to £4m has been put aside to support the creation of IRs in the UK

‘grey route’ - Publisher-led experiments – for example:

•Springer – Open Choice: author chooses to either pay \$2,000 for OA or traditional publishing route with no pre-payment; Blackwell Publishing – Online Publishing (\$2,5000 per article); OUP – experimenting with OA for some learned society journals; Nature – allowing authors to archive on IR after 6 months; US learned societies – digitising backruns of journals and making freely available through HighWire platform. Costs being absorbed by current subscription and license model. *Currently this represents the largest mass of ‘open access’ scholarly information available*

Changing user needs, expectations and behaviour

- Greater specificity, personalisation, comprehensiveness and speed / ease of access
- Content alone decreasing in value; growing demand for subject based, fully connected content of all types, including 'new/ non-traditional', integrated with software and adapted to workflow
- Increasingly expect not to have to pay

But

- Still the need to retain:
 - Certification = quality
 - Relevance = collection
 - Browsing, searching and finding = navigation
 - Persistence and continuity = historical files / archives

UK developments

- **Science & Technology Select Committee Inquiry into Scientific Publications**
- **Research Councils UK (RCUK) policy on the publication of and access to research outputs**
- **Office of Science and Technology (OST) e-infrastructure group – assessing current state / future requirements of the UK's e-infrastructure**
- **Joint Information Systems Committee (JISC) – supporting projects and programmes to test open access models**

These are some of the major developments which have been taking place in the UK. I'll come on to talk about each of these in more detail.

Science and Technology Committee Inquiry: took place last year and reported in July 2004. Inquiry was looking at access to journals within the scientific community, with particular reference to price and availability. Many organisations – including the BL – provided input to the inquiry.

RCUK acts as a representative body of the seven UK Research Councils (which together spend over £2 billion each year on research). RCUK is about to publish its policy for access to research literature.

OST e-infrastructure group was set up in response to recommendations arising from the Government's 10-year Science & Innovation Investment Framework, which was also published in July 2004 (I will come on to talk about this in more detail later)

JISC (Joint Information Systems Committee) provides strategic guidance, advice and opportunities to use ICT to support teaching, learning, research and administration in higher and further education in the UK. It has been a supporter of open access: for example, supporting the FAIR programme for institutional repositories and the ROMEIO list of publisher copyright policies, and providing transition funding for open access journals. JISC provides funding for a number of projects in this area – will come on to discuss this in more detail a bit later on.

Select Committee Inquiry into Scientific Publications - conclusions

- **Concluded (July 2004) that the current model for scientific publishing is unsatisfactory for both libraries and users**
- **Put forward recommendations which it felt would increase access to the literature in the UK, including:**
 - the establishment of institutional repositories in all UK higher education institutions for storage of and free access to their published output
 - mandate from Research Councils and other Government funders that all funded researchers deposit a copy of all their articles this way
 - further experimentation and testing of open access models, helped by the establishment of a Research Council fund to support funded researchers should they wish to pay to publish
- **Recommended that Government formulate a strategy for future action as a matter of urgency**

Select Committee Inquiry – recommendations of particular interest to the BL

- The British Library's Document Supply Service is an efficient and cost-effective method of providing access to articles in scientific journals. The decline in demand for Document Supply notwithstanding, we are persuaded that the service provides a valuable alternative route for users who would not otherwise have access to the journals that they needed. We recommend that the Government takes steps to protect the service. (Paragraph 31)
- We recommend that DCMS provide adequate funds for the British Library to establish and maintain a central online repository for all UK research articles that are not housed in other institutional repositories. (Paragraph 118)
- The British Library has a crucial role to play in the preservation of digital publications, both strategically and practically. This is an expensive process. Whilst the publication of this Report is too late to have any influence on funding decisions made as part of the 2004 Spending Review, we strongly support the British Library's call for extra funding in recognition of the work that it has carried out in this capacity. Failure of the Government to give adequate funding to the British Library could result in the loss of a substantial proportion of the UK's scientific record. (Paragraph 196)
- Gaps of up to 60% in the deposit of electronically-delivered publications, including STM journals, represent a significant breach in the intellectual record. It is imperative that work on recovering and purchasing the missing items begins immediately. The six deposit libraries will need additional funding to do this. (Paragraph 203)
- It is vital that work on regulations for the legal deposit of non-print publications begins as soon as possible. (Paragraph 199)
- The existence of a secure network between the legal deposit libraries would create greater efficiencies in the deposit system and would have the potential to increase access to deposited material. We recommend that provisions for such a network are made in the regulations with these two aims in mind. The deposit libraries should be funded to establish the network. (Paragraph 201)

Here we have pulled out just a few of the recommendations which are of particular relevance to BL

First two bullet points relate to the BL's role in underpinning research in the UK by maintaining a central, accessible repository of research information. Recognises the value the BL provides by enabling access to research articles. We are continuing to provide remote access to articles held in our collection and to improve the way we do this in line with user needs (document supply & SED – will talk more about SED later on)

- SED
- Personal use
- New business models

Third bullet relates to the BL's role in developing the national digital repository and the critical need for funding. This was a central part of the BL's bid for funding in SR04, and we have continued to push this message – e.g. in our response to the Govt's 10-year framework for science. Work being taken forward in this area by OST e-infrastructure group (will come onto discuss)

Fourth & fifth bullets relate to e-legal deposit, and the need to ensure that the full UK published archive – whatever the format – is captured and made accessible over the long term. I will talk later about how the Legal Deposit Libraries Act 2003 is being taken forward.

Select Committee Inquiry – Government's response

Disappointing response from UK Government

- **Wants to see the outcomes of publicly funded research made available to widest possible audience, providing that:**
 - quality maintained
 - access provided at reasonable cost
- **Recognises potential benefits of IR and welcomes increasing number of Universities providing these to increase access to research output**

But...

- **Not directly taking up the recommendations**
- **'Not aware that there are major problems in accessing scientific information'**
- **Not providing extra funding, e.g. for the development of institutional repositories**
- **No mandate to Research Council funded researchers to deposit with institutional repositories**
- **Remains to be convinced that the author-pays model is better and cheaper**

The Inquiry had concluded by making a number of recommendations which it felt would increase access to the literature in the UK. It was therefore very disappointing that the Government did not directly take up or adopt any of these recommendations.

Research Councils UK

- **Welcomed the Select Committee's Report on Scientific Publications**
- **Has developed a cross-Research Council policy on publication of and access to research outputs, which provides a fuller view on many of the issues raised in the Select Committee report**
- **Not yet published but we expect the following:**
 - Emphasis on institutional repositories rather than open access
 - Strong emphasis on peer review
 - RCUK requirement that, from October 2005, grant holders place publications and conference papers in a repository (institutional or subject), where such a repository exists
 - Deposit to be at earliest opportunity
 - System of deposit to be reviewed by the end of 2008

RCUK is expected to publish its policy imminently.

OST e-infrastructure group

- **Established in response to recommendations arising from the Government's 10-year Science & Innovation Investment Framework (July 2004)**

"The growing UK research base must have ready and efficient access to information of all kinds...This is the life blood of research and innovation.

The Government will therefore work with interested funders and stakeholders to consider the national e-infrastructure (hardware, networks, communications technology) necessary to deliver an effective system. These funders and stakeholders include the British Library, which plays an important role in supporting scientific research and potential, including providing benefits to smaller business in the UK through access to science, engineering and technology information sources"

Science & Innovation investment framework 2004-2014

The OST e-infrastructure group was set up in response to recommendations in the Government's 10 year science and innovation investment framework.

Owing in large part to BL efforts, the 10-year framework includes 3 paragraphs about 'Information Infrastructure' (2.23-2.25) which incorporate an action for Government to:

'work with interested funders and stakeholders to consider the national e-infrastructure (hardware, networks, communications technology) necessary to deliver an effective system'

OST e-infrastructure group

- **Cross-departmental representation:** BL, Cabinet Office, HEFCE, JISC, RCUK, DfES, DCMS, OST
- **Remit is to:**
 - take a serious, coherent look – from researcher's perspective - at the UK's e-infrastructure
 - identify strengths & weaknesses in light of future requirements
- **Resulting shared diagnostic 'roadmap' to be used to identify priorities for investment going forward – in particular, to support bids to Spending Review 2006**
- **BL joint-funding consultancy work to develop the 'roadmap'**

BL is playing an active part on the OST e-infrastructure group

We are joint funding the piece of consultancy work with JISC and eScience to develop the roadmap

The roadmap will provide an overview of current e-infrastructure provision in the UK and will compare this with other countries.

The roadmap will be used as a fact base to identify strengths and weaknesses in the UK's current provision, which ultimately will lead to recommendations. It is anticipated that these will be used to lobby government to recognise the importance of e-infrastructure to UK research and innovation.

JISC – relevant initiatives in which the British Library is involved (1)

Supporting a range of projects to test and develop open access models

SHERPA

- Aims to set up institutional open access e-print repositories in 20 partner research library institutions
- BL's role is to establish a repository for independent, non-institutional researchers

E-theses national pilot

- Purpose is to deliver a fully operational, scalable, financially viable prototype UK e-theses online service and infrastructure
- Builds on the national theses service currently provided by the BL on microfilm
- Will investigate mass digitization of current holdings in the BL and universities

Preservation e-print services

- Aim is to implement an ingest service based on OAIS reference model for institutional archives
- BL & Southampton University will build and test and exemplar OAI-based preservation service
- Service could be used with any OIA-compatible preservation archive to create a software-independent archive

SEE NEXT PAGE FOR SPEAKER NOTES

JISC – relevant initiatives in which the British Library is involved (2)

Virtual research environment

- Looking at embedding a VRE in an institutional environment
- Project will test the integration and deployment of key existing software components within a portal framework at Leeds University
- BL is testing delivery of BL services from within a local institutional VRE and assessing potential long-term preservation requirements for materials held in a VRE

Lifecycle information for e-literature

- Project will explore and develop a lifecycle approach to costing digital archiving for e-journals
- Project outcomes are intended to answer key questions for HE & FE such as:
 - What are the long-term costs of preserving digital material?
 - Who will do it?
 - What are the costs for a library in HE/FE to partner with another institution to carry out long-term archiving?
 - What are the relative risks of digital vs. paper archiving?

SEE NEXT PAGE FOR SPEAKER NOTES

British Library initiatives

- E-infrastructure development
- E-legal deposit
- Web archiving
- Digital archiving services
- Digital access to research articles

E-Infrastructure development – Digital Object Management Programme (DOM)

Mission

To enable the United Kingdom to preserve and use its digital intellectual property forever

Vision

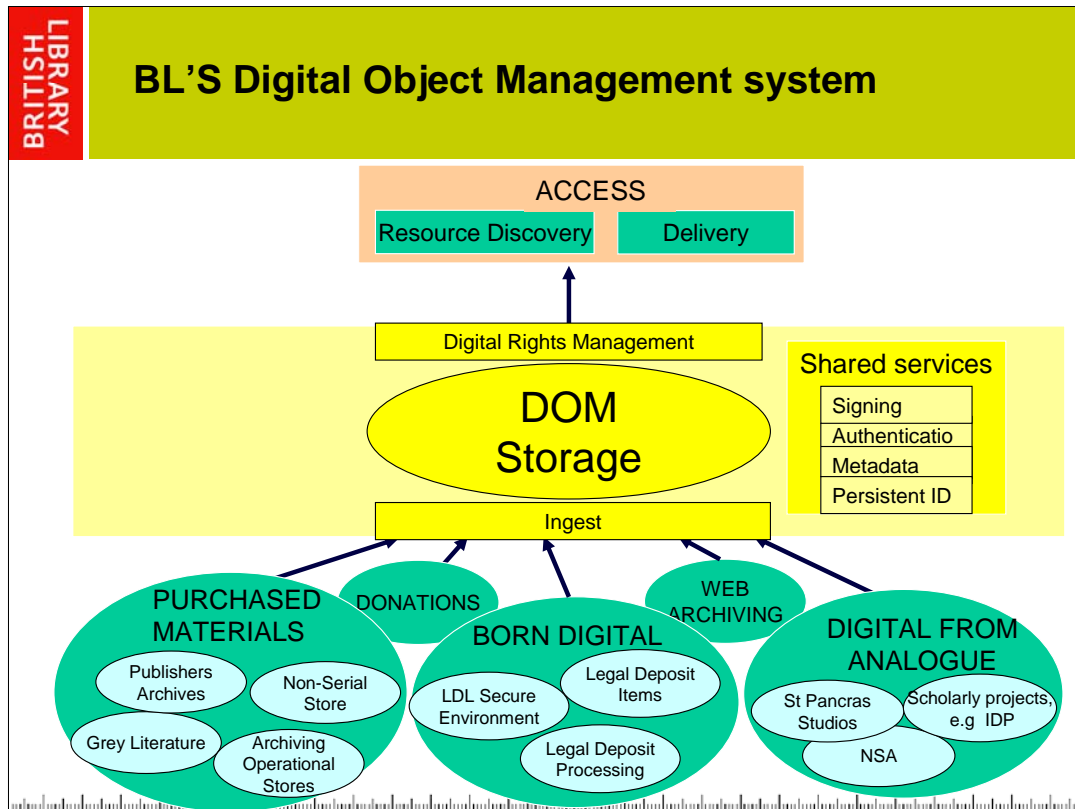
To create a management system for digital objects that will:

- store and preserve any type of digital material in perpetuity
- provide access to this material to users with appropriate permissions
- ensure that the material is easy to find
- ensure that users can view the material with contemporary applications
- ensure that users can, where possible, experience material with the original look-and-feel

Drivers for the DOM programme:

- Legal deposit legislation for non-print material was granted royal assent in October 2003
- Existing voluntary deposit scheme operational since 2000
- Storage of digitised masters from early '90s onwards
- New digitisation initiatives: newspapers, sound, etc
- Sound archive receives 12T of material per year (with 50 year collection)
- Web archiving
- Cartography and datasets
- Electronic journals, picture library
- ... and
- and

We need a generic and cost-effective approach for the secure long term storage of digital material that is produced by numerous initiatives



Work well underway on designing and developing the core storage functionality

The system architecture and incremental approach to development will enable us to build on this preservation-quality digital store by delivering additional services through subsequent, regular releases.

Key design requirement is that the system should be able to mix storage hardware from different manufacturers, permitting procurement every few months and simplifying the replacement over time of older units.

A second requirement, achieving an acceptable level of disaster tolerance within a reasonable budget, has led to the adoption of a multi cluster design.

Recent change in direction due to performance issues with a key infrastructural component from Microsoft. A re-design has been undertaken which will result in a far more performant and scalable core to the system

First release of the storage system will provide a preservation-quality digital store for material received under Voluntary Deposit scheme. Second release will extend this service to the digitised master files held by the Library

Technical direction of the solution architecture has been validated by two external Technical Advisory Panels.

E-legal deposit – defining the legislation

Legal Deposit Libraries Act 2003

- Provides for deposit of electronic publications
- Secondary legislation in the form of Regulations is needed in order to implement the Act

Legal Deposit Advisory Panel

- Being established to advise the Secretary of State on the timing and content of Regulations
- Panel will consist of a Chair and 14 members
 - 5 members to be from publishing industry
 - 5 members to be from legal deposit libraries
 - 4 members to be from independent fields with relevant expertise
 - Chair will be appointed by the Secretary of State
- Alongside this, publishers and libraries will also continue to work together via the JCLD

The aim is to create a panel that has an understanding of the wide range of issues arising out of the legal deposit of non-print material, and can provide fair and balanced advice on the incremental implementation of the Legal Deposit Libraries Act 2003

The remit of the Panel is purely advisory - no formal responsibility for the administration of the 2003 Act

A consultation document on the roles / responsibilities of the Panel was published in November 2004 on the DCMS website. Public consultation ended 16th February

Deadline for applications is mid April

Panel should be in place by the Autumn (first regulations not likely to be in place before 2006)

These groups will hold ex-officio seats on the Advisory Panel:

- BL
- National Library of Scotland
- National Library of Wales
- Digital Content Forum

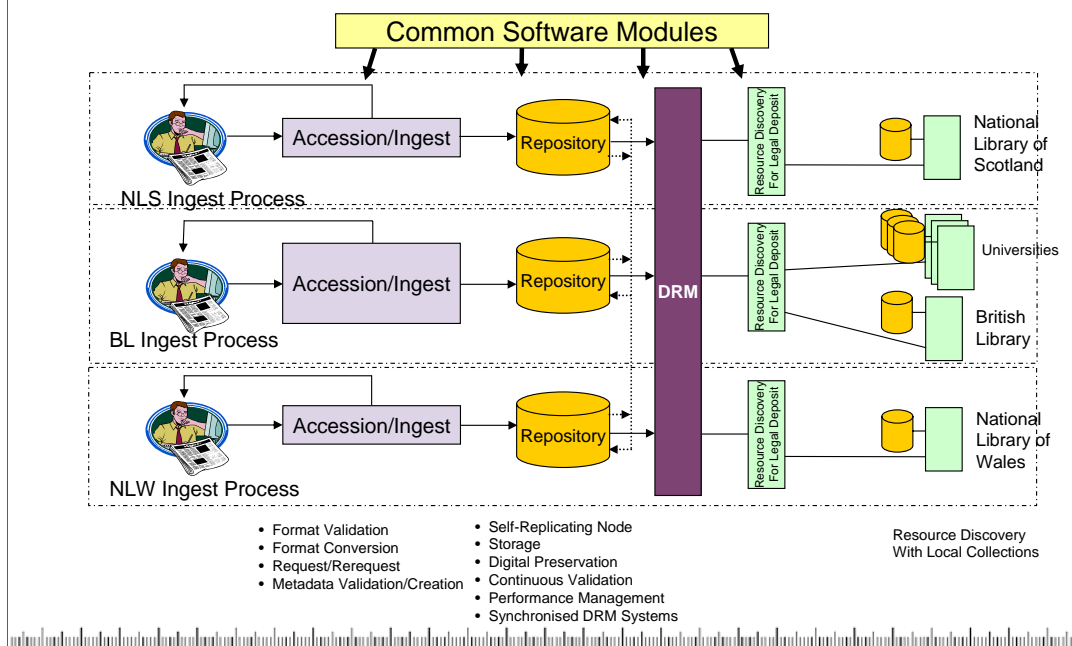
JCLD (Joint Committee for Legal Deposit)

Three main pieces of work being taken forward:

- Territoriality (defining place of publication)
- E-journals pilot scheme (will discuss in more detail in a moment)
- Offline / handheld publications (CD roms, microfilm, etc)

Deposit of offline publications will probably be the first to be regulated

E-legal Deposit: possible architecture



The BL has been leading the Legal Deposit Libraries Committee working group on digital infrastructure.

This has been looking at the options for a digital infrastructure that will be secure and cost-effective, while meeting the needs of each of the legal deposit libraries for deposit, storage and access.

This diagram shows a possible architecture. Discussions still ongoing with other legal deposit libraries regarding the final architecture

E-legal deposit – e-journals pilot

Aims of the e-journals pilot

- to test the technical infrastructure, mechanisms and procedures relating to the deposit, storage and preservation of electronic journals
- to highlight any interface problems, facilitating their early resolution
- to produce a fully operational and scalable means for the deposit of e-journals, which can then support access models as developed and agreed subsequently
- Pilot will run for up to 12 months from April 2005

Who is participating

- Pilot developed by a working group under Joint Committee on Legal Deposit
- Publishers have volunteered (via their trade associations) a sample of e-journals offering diversity of subjects and formats
- c. 30 publishers involved

In anticipation of Regulations to further define the terms of the Legal Deposit Libraries Act 2003 we are participating in a pilot scheme to test infrastructure and processes.

c. 30 publishers participating (from small publishers, e.g. learned societies, through to large commercial publishers)

A prototype e-journal metadata deposit tool is being developed by the BL to allow ingest of the range of non-uniform data

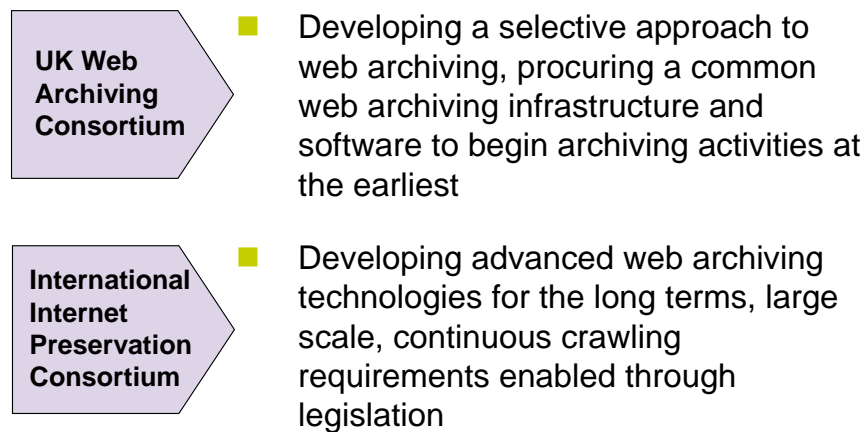
Although no end-user access is permitted under the pilot scheme, the libraries will have appropriate access rights to manage the material in their digital asset systems.

Ultimately, Regulations will provide for a limited amount of public access within the libraries.

Therefore, in parallel to this technical pilot, the working group will separately begin to consider what the appropriate level of access might be, using the access regime for the offline voluntary scheme as a starting-point for their deliberations

Web Archiving - structure of BL's programme

The BL's Web Archiving Programme is a collaborative initiative, roughly implemented across two consortiums



We are collaborating with national and international partners on our web archiving programme

We are actively archiving in an experimental capacity to build experience, test new technologies and begin the process of capacity planning, scalability planning, crawl scheduling, etc.

We have an engineering team in place, working on archival web crawlers, archiving access tools and scheduling systems

We also have a web archive curator working on the collection policy and subject / site selections

UK Web Archiving Consortium

- Developing a selective approach to web archiving
- License for PANDAS signed with NLA
- Sub-licenses with consortium partners and contractor to follow
- ITT concluded with Magus Research winning the contract.
 - Implement a common web archiving infrastructure (lots of Linux machines + PANDAS)
 - Provide customisation/development of PANDAS
 - Provide help desk and support

BL is leading a UK consortium on a selective permissions-base project to collecting selected UK web sites.

We have managed to set up web archiving here and within UKWAC thanks to PANDAS and the NLA. We recognize that PANDAS has faults and are grateful to NLA for waiving licence costs to UKWAC. PANDAS is likely to be replaced by curator tools being developed under the IIPC

Web Archiving is now taking place here. We have selected 493 sites and to date the BL has archived c. 60 sites including: UK-related tsunami sites; London bid for the Olympics website; National Museums Directors Conference; Hindu Forum of Britain; Crisis (charity & campaigning organisation) We are about to devote attention to sites around forthcoming general election.

Consortium as a whole has so far archived 225 sites

Intention is for UKWAC site to go live by early June - we don't want to risk launch at time of intensive activity around gathering, archiving for election

International Internet Preservation Consortium

Developing advanced web archiving technologies:

- Smart Crawler
 - Continuous adaptive crawler, adjusting crawl priority on the fly
 - Based on IA Heritrix
 - Working on requirements now
 - Expect to being tender process in June
- Content Management
- Archival formats
- Framework
- Metrics and Test Bed

The BL is part of an international consortium to develop web archiving technologies to meet our long term needs (i.e. national domain level archiving, to enable the BL to collect occasional snapshots of the entire UK web, with selected more frequent archives of sites of importance)

Digital Archiving: the British Library's proposition

- The British Library is committed to delivering a long-term digital architecture which will preserve the UK's electronic content in perpetuity – with a key focus on legal deposit material
- The British Library is also working on services and business models to ensure that electronic content is also available to subscribers in the event that publisher content becoming unavailable (e.g. through system failure or company failure). This will give publishers and libraries far more confidence in the switch to digital-only content
- The British Library already provides 'backup' services of this kind. Our electronic storage system already contains back-up copies of ten of the largest STM publishers – and we are working with other publishers to obtain more content
- The biggest bottleneck to developing this is funding. We have yet to receive any government funding for this key initiative, although OST has indicated support in principle. We would like to escalate this development in the interests of UK researchers, libraries and publishers

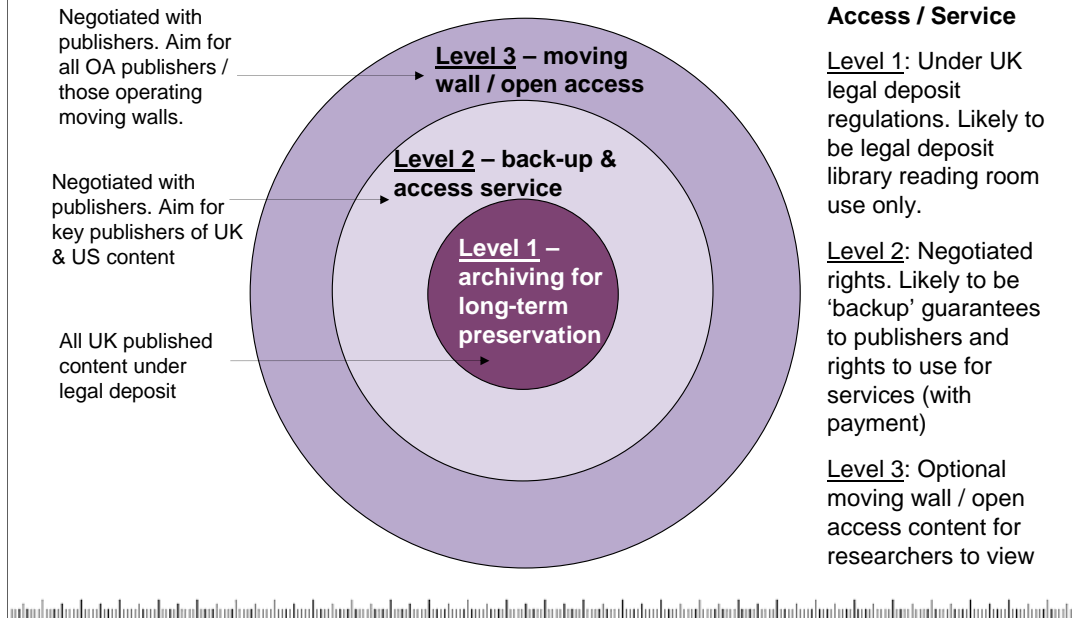
We recognise that our DOM infrastructure (with the agreement of contributing publishers) could be used for the provision of additional value added services

We are currently working on potential business models which might extend the levels of access or content beyond that available under the narrow remit of UK legal deposit legislation

Because of commercial confidence issues these plans have not as yet been widely discussed or disseminated

We believe that a service which offered levels of 'back up' and guaranteed long term access would both offer security to the library community, allowing the switch from paper to 'e' content (with associated efficiency gains) and also allow the BL to play a wider service role with potential new income streams.

A possible digital archiving service model



Our initial thoughts on the service model required are based on the idea of 'levels' of rights and services as shown here.

Improving access at article level – Secure Electronic Delivery and British Library Direct

Secure Electronic Delivery

- Launched in 2003
- Access to over 100 million items
- Uses Adobe Reader 6 (freely downloadable) to deliver electronic documents securely to customers' desktops
- Fully copyright compliant
- Rapid delivery to desktop, within 2 hours if required

British Library Direct

- Launching in April 2005
- Article-level access to a database of our 20,000 most requested titles
- 5 year back file on all titles
- Pay-as-you-go: no registration or subscription required
- End users can search the database for free and order articles using their credit card
- All articles available for electronic delivery in as little as 2 hours - choice of delivery times

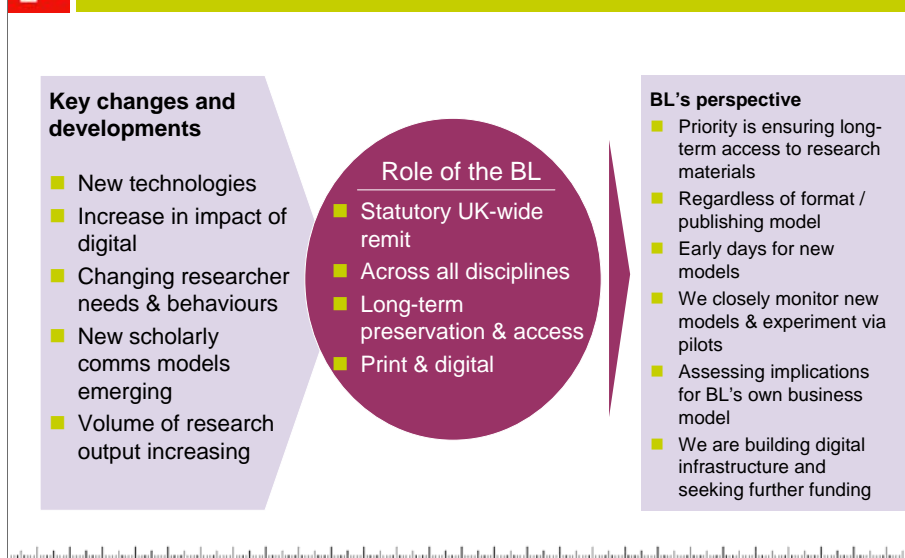
SED

- Secure Electronic Delivery (SED), was launched at the International Online Meeting at Olympia in December 2003. This service gives secure access to 100 million documents which can be delivered electronically to customers' own PCs
- It represents a major output from the ISB-funded document supply modernisation programme.
- We worked with individual publishers to agree licences to hold electronic content and developed a service, using Adobe products
- We are the only document supplier offering secure electronic delivery on this scale

BL Direct

- Users will get access to bibliographic records of key journals within their field AND outside their main field of interest. It will include medicine, pharmacy, engineering, pure science, food and agriculture, economics, education, environment, law, industry
- No registration or subscription required to search the service. Users can just 'grab and go'. (with SED, users have to register)
- It serves organisations of all sizes. Smaller organisations can access articles without having to pay a subscription fee and larger organisations can use it to support their existing subscriptions.

Summary – the BL's perspective



- The efficient and effective flow of data, information and ideas is critical to the health of UK research and innovation.
- Research libraries – and the BL in particular – play a crucial role in the UK's information infrastructure, preserving and making accessible the information needed to support research.
- Digital technology is transforming the nature of research. This means that the environment in which the BL operates is becoming increasingly complex.
- New challenges arise from this complexity – e.g. how to preserve digital formats; how to prevent obsolescence when technology develops; how to fund the development of new infrastructures, systems and skills to cope with digital information
- However, the BL's core role does not change. Our priority remains - collecting and providing long-term preservation and access in relation to the UK published archive.
- In terms of the new models such as open access publishing we are still in the very early days. For example there are 1513 titles registered in the Directory of Open Access Journals –the BL currently subscribes to c.35,000 journals.
- Our key concern with any new models is that long-term preservation and access issues are addressed. These models may make information accessible now in a convenient way, which is less costly at the point of use, but how will the information be archived for the long-term?
- The BL has an important role to play in incorporating the various new forms of publishing into a national information infrastructure. The BL will also continue to play a critical role in navigation – just because materials is 'out there' and freely available doesn't make it any easier to find.
- We are developing our own capabilities in terms of digital infrastructure (national repository) and our skills
- We are actively experimenting in new models via pilots (e.g. JISC projects)
- The implications for our costs are not clear cut – but it is unlikely that they would go down (and there are potentially big financial impacts on our revenue further down the line if we do move to open access, so at the BL we are also assessing these implications). It is vital that we consider the economics of new models in a life-cycle and service context.
- Developing the infrastructure to support long-term preservation of digital research materials is a significant task. We are therefore seeking funding from Government to help us.