

eJournals at the British Library National Action via Legal Deposit

**Dr Lee-Ann Coleman
Head of Science, Technology & Medicine**

**26 April 2010
DPC: eJournals are forever?**

The British Library:

‘This is the life blood of research and innovation’

Science and Innovation Investment Framework 2004-2014, H.M. Treasury (2004)

Information infrastructure

2.23 The growing UK research base must have ready and efficient access to information of all kinds – such as experimental data sets, **journals**, theses, conference proceedings and patents. ***This is the life blood of research and innovation.***

National library of the UK and a legal deposit library.
Serve researchers, business, libraries, education & the general public

Collection includes over 30k journal titles, 5m reports, theses and conference papers

Print collection fills over 600km of shelving, grows at 11km per year
70 Tb of digital material via voluntary deposit & digitisation projects

Business and IP Centre:
Providing inspiration, and enabling protection of creative capital and business development



**Helping people
advance knowledge to
enrich lives**

The largest document supply service in the world. Secure e-delivery and ‘just in time’ digitisation enables desktop delivery within 2 hours

Generates value to the UK economy each year of 4.4 times public funding

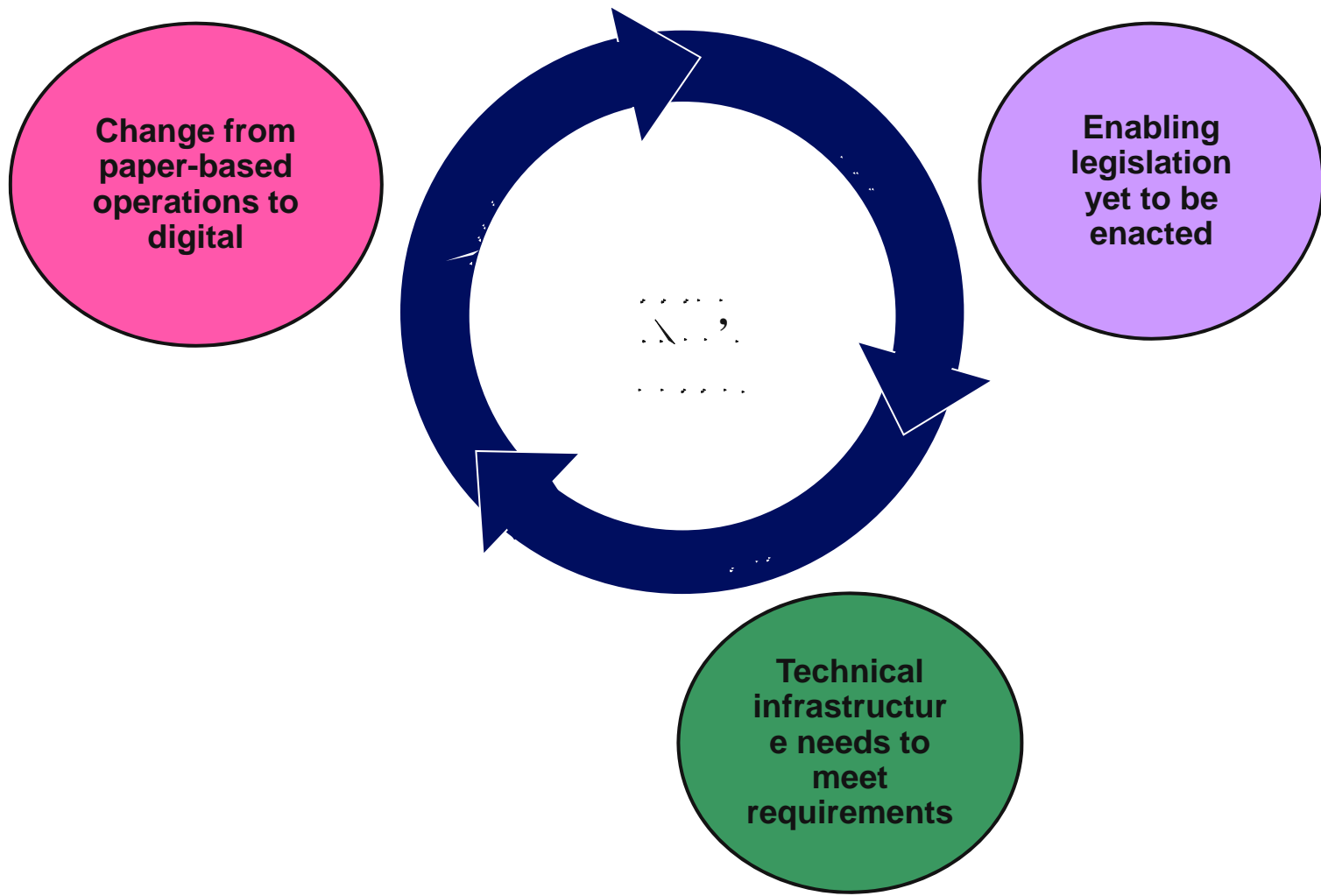
GIA Funding 08/09:
£94.8m operational,
£12m capital

Other funding secured 07/08:
c.£33m

3 main sites in London and Yorkshire. Circa 2,000 staff

Preparing for digital

for everyone who wants to do research

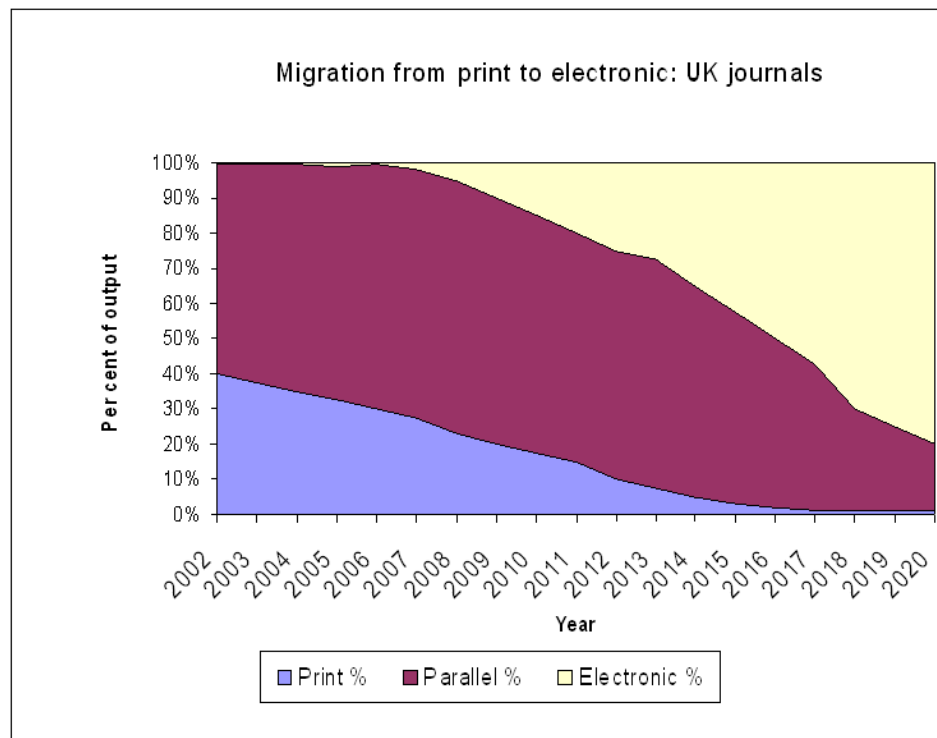


Legal Deposit Libraries Act 2003

Essentially an update of 1911 Copyright Act

Primary legislation which

- 1) **reaffirms print deposit**
 - 2) **creates a framework for secondary legislation to cover non-print deposit**
- Regulations need 'affirmative resolution'
 - No distinction between the libraries
 - Access restricted to library premises
 - Excludes sound and film where they are the primary content
 - Publications issued in both print and digital media



Currently, publishers are not obliged to provide us with electronic copies

Gathering information (a 2008 'snap shot')

Study commissioned from Rightscom Ltd

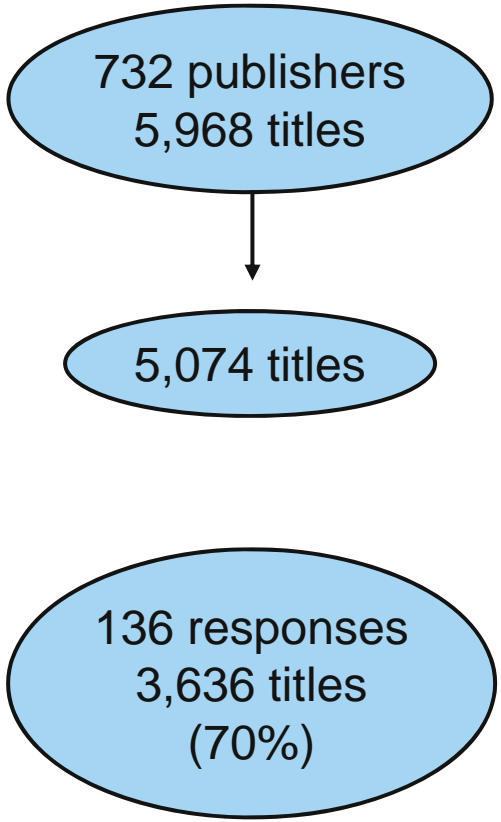
eJournals prioritised

Ulrich's Directory of Serials

ISSN Registry

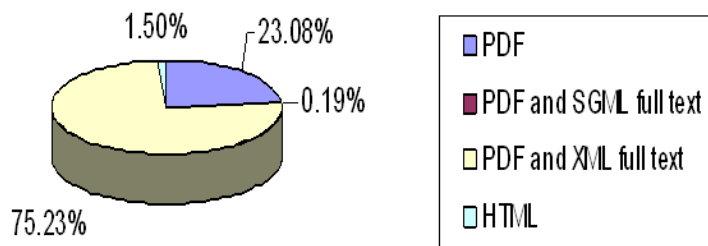
De-duplication etc

Questionnaires (608 publishers)

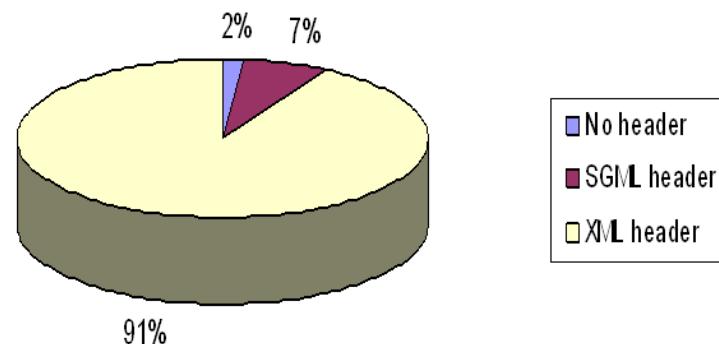


Summary of Rightscom Findings - File Formats

**Main content file formats, % of titles
(n=3609)**



**Metadata file formats, % of titles
(n=3710)**



- Little use of SGML, with the majority of titles in PDF & XML full text (strong trend to NLM publishing as ext. DTD)
- A significant minority in PDF only and some in HTML only
- Supplementary files in a wide variety of formats
- Polarisation between XML headers & no metadata, or metadata supplied via hosting services.
- SGML headers are disappearing.

Interpreting the eJournals Landscape

Large Publishers

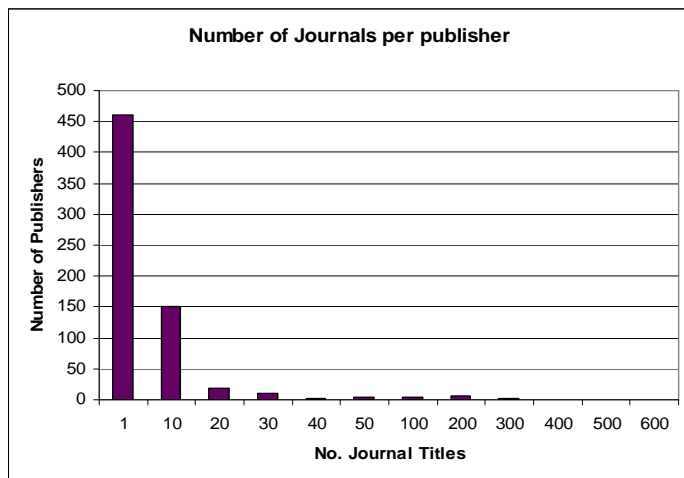
- High volume but small number
- Multiple options for target audiences
- Small no. of DTDs within organization
- Accompanying metadata more common

Mid sized publishers

- Lower volume
- Fewer options for users based around PDF with some XML
- Minimal metadata

Small Publishers

- Large numbers
- Low volume
- PDF/HTML/Word
- Plus some proprietary solutions
- No metadata
- Irregular publication



460 publishers: 1 title

150 publishers: 2-10 titles

52 publishers: >10 titles

The Digital Library Programme

Responsible for providing the hardware infrastructure and associated software tools to **acquire, process, store, preserve, and make accessible** digital items.

Content stream	Count	Volume(GB)
Sound Archive	60,110	9,593.71
VDEP	920,982	3,990.27
eJournals Stream 2	5,338	14.37
Legacy Newspapers – JISC1	1,052,833	33,328.93
19th Century Books	322,409	22,625.59
Born Digital Newspapers	7,119	166.72
Total	2,368,791	69,719.59

The Digital Library System

■ Scalable

- Multi-server Architecture
- Storage vendor agnostic

■ Secure

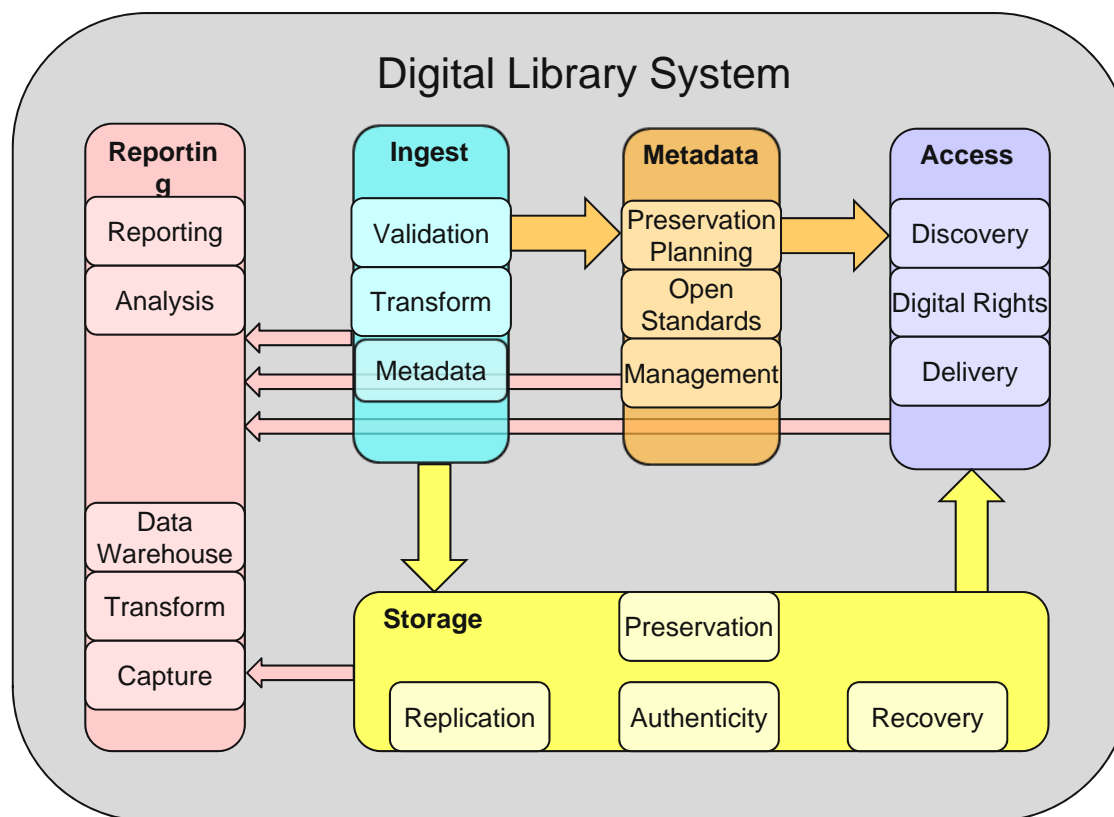
- Layered Security Model
- No trust between sites

■ Resilient

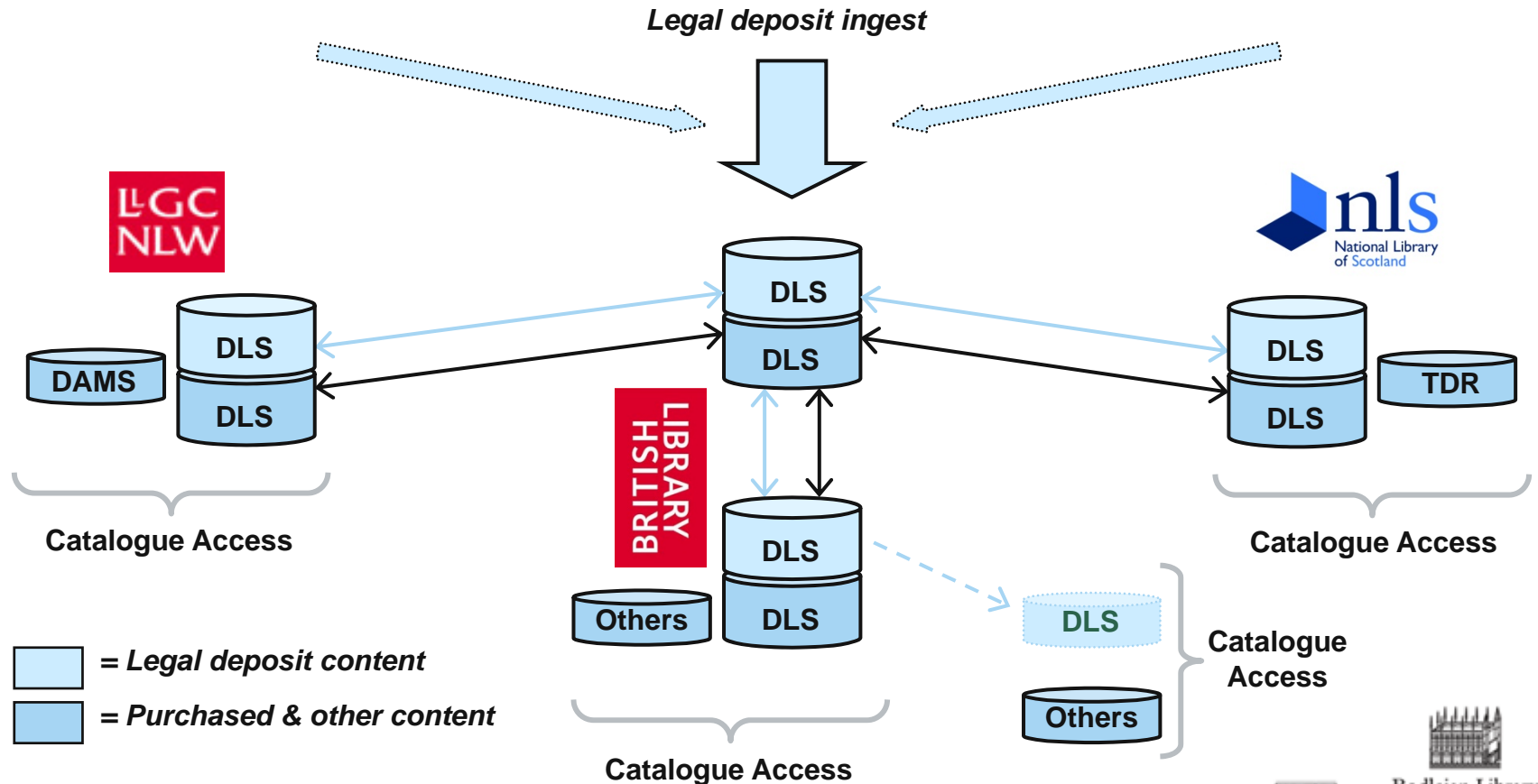
- Self healing
- Multiple copies

■ Provide Provenance

- Digital Certificates
- Secure Timestamps



Tripartite Shared Infrastructure for Legal Deposit



DAMS = Digital Asset Management System
DLS = Digital Library System
TDR = Trusted Digital Repository

Working with eJournal Publishers to date

- Requirement to test voluntary scheme and resource requirements
- Pilot project conducted in 2006
- **Full text ingest** (with transformation) mechanism:
 - Only publishers able to meet specific technical criteria invited
 - Full-text XML to the NLM A&I DTD v2.3 as well as PDF v1.7 format using FTP as the file transfer protocol
 - 16 publishers contacted - 9 agreed to deposit voluntarily
 - In October 2009, ingest into live store unsuccessful using software developed by 3rd party suppliers
 - Since that time reviewing options
- **PDF and Word ingest** mechanism:
 - Iterative software development solution (in-house)
 - Initial tests ingesting issue level material from 100 publishers
 - ~200 small to medium publishers invited to submit PDF or Word docs
 - Around 28 participating; ingest at article level achieved
 - Developing operational capacity

Access: Only one user at a time, on-site, in each deposit library

The future?

- Legislation
- Internal readiness and solutions for ingest
- Continued close working with publishing community
- Preservation and access of outputs of UK scholarly publishers
- Building an important resource for the future