Defining National Solutions for Managing Book Collections and Improving Digital Access

Neil Grindley, Head of Resource Discovery, Jisc
Paola Marchionni, Head of Digital Resources for Teaching, Learning and Research, Jisc
The National Monograph Strategy
2012 - 2014

The Vision

Within 5 years UK researchers and students will have unparalleled access to a distributed national research collection enabled by an open collaborative national infrastructure.
Issues of concern to UK academic libraries
• Management of print collections (space management)
• Collection management (benchmarking and development)
• Collaboration to reduce duplication (data, cataloguing effort, resource investment)
• Poor data quality
• Disambiguation and de-duplication of metadata records
• Discovery of digital surrogates
• The procurement, management and discovery of e-books
• Metadata standards and persistent identifiers
• Licensing, rights and permissions for access
• National and international authority sources
• National coverage
• Building sustainable and effective platforms
• International engagement
• Data formats and emerging opportunities for exploiting data
• Preservation and conservation
• Consuming the data and the end user perspective

(source: Jisc Bibliographic Data Roadmap Group: September 2014 - January 2015)
What problems are we trying to solve?

Libraries want to make data-driven decisions about the management of their print and digital book collections but the data that is currently available does not allow them to do this with confidence.

Libraries want to ensure that researchers and learners have sustainable and convenient access to digital books but it is currently not obvious what is available or what could readily be made available.
Principles …

1. The UK has a fundamental need for a new national-scale service to drive a range of required functions.

2. The new service should consist of an aggregated database and its management should be outsourced to an organisation that is capable of delivering the service as core business at scale.

3. The primary focus of future effort should be on supporting UK academic libraries with collections management. Resource discovery and records delivery are of secondary importance.

Principles ...

4. The data contributed to the new system must remain sharable and reusable by all contributing organisations and by other relevant organisations that support discovery and records delivery.

5. The route to greater impact for contributed library data is through exposure to global search engines and other high impact web-scale channels rather than through reliance on Jisc-funded discovery interfaces.

6. The new system should combine knowledge about both print and digital publications for services to be efficient and effective.

The National Bibliographic Knowledgebase (NBK)

Bibliographic Data
- Legal Deposit Libraries
- Academic Research Libraries
- Teaching & Business Engaged Libraries
- Specialist Libraries
- Open Access Book Directories

Availability Data
- Identifiers
- Standards

Usage Data

Supported Functions
- Benchmarking Collections
- Managing Print Collections
- Document Delivery
- Digitisation and Preservation
- Resource Discovery
- Copy Cataloguing

Supported Functions
- Benchmarking Collections
- Managing Print Collections
- Document Delivery
- Digitisation and Preservation
- Resource Discovery
- Copy Cataloguing
The National Bibliographic Knowledgebase (NBK)

Identifiers

Standards

Availability Data

From a library
From a publisher
From the Web
From a repository

Usage Data

Circulation data
Usage data
Demand information

The National Bibliographic Knowledgebase (NBK)
The National Bibliographic Knowledgebase

Support services
- Community Engagement
- Advice & guidance
- Data quality enhancement

Identifiers

Standards

Tools & services layer
- Collection Management
- Data Sharing
- Metadata Linking & Research
- Version tracking & linking
- Support for new publishing models

The National Bibliographic Knowledgebase
NMS report recommendations:

Rec #5: Digitisation strategy

Building a national digital research collection

Rec #6: Licencing strategy

A nationally negotiated agreement for digital monographs

Two sides of the same coin:
- Ambition to increase access to digital monographs as part of a national digital collection
- Of benefit to collections managers to enable more informed collections management decisions
- Based on evidence of need
Where to start?
- Public domain?
- Orphan works?
- In copyright?

What would justify investment and ‘buy-in’?
- In print?
- Out of commerce?
- Discipline specific?

What publishers offer?
We asked libraries where the market failure was. Which books did they want digital access to and why couldn’t they get it?

- Durham University
- Royal Conservatoire of Scotland
- University of the Arts
- University of East London
- University of Glasgow
- University of Manchester
- University of Portsmouth
- University of St Andrews
- University of Sussex
- University of York

Institutional pilots to solicit “problem” titles, identify barriers and explore solutions Jan-Jul 2016
A wish-list of titles (1117 sample)

<table>
<thead>
<tr>
<th>Titles</th>
<th>Usage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Portugal. Ministerio dos Negocios Estrangeiros. (1973). Portuguese Africa: an introduction. Lisbon: The Ministry</td>
<td>Chapter requested as part of a reading list for undergraduate course. CLA don't have an agreement with Portugal. The fact that it is a government publication does not make a difference</td>
</tr>
<tr>
<td>Bacon, Francis, (ed.). (1984). Francis Bacon's natural philosophy: a new source: a transcription of manuscript Hardwick 72A. Chalfont suitable repository for the content so decision made to print hard copy for library collection. Constantly updating publication so digital would have been far more efficient and accessible.</td>
<td>Book requested for an undergraduate teaching course. E-book not available to purchase</td>
</tr>
</tbody>
</table>
What we found out

• Work in progress: data is still being tided-up, uniformised and triangulated with Nielsen database, more difficult than it sounds
• This is just a sample so will have a degree of bias, but we are looking for patterns
• Especially publishers’ names are likely to be different following mergers
Availability problems by titles (1117 sample)

- Majority of titles are Print-only
- A good proportion of e-books are still “unavailable” to libraries even if an e version exists (reasons eg: non available in UK, no inst licence or unsuitable to needs, too expensive; not on right platform)
- Most titles seem to fall within arts & humanities and social sciences, but no “scientific” analysis done
• Majority of in-demand titles were published from the 1980s onward
• No interest in Public Domain books
Top 10 publishers (1117 sample)

- 20.4% of titles (228) are Print-only/out of commerce
- Publishers and titles distribution by category will change as some publishers have merged with others, we’re checking this
- Long tail
Users

- Single user
- 2-5 concurrent
- 5-50 concurrent
- >50 concurrent

Libraries’ requirements

Need

- Preservation
- Reading list
- Research
Requirements that a solution should satisfy

1. Ability to “aggregate” problem titles from libraries’ reading lists
2. Check reliability of bibliographic data (publishers, ISBNs..) from libraries against authority source (eg Nielsen database?)
3. Identify copyright owners and obtain licencing permissions
4. Cater for a “long tail” pattern of requests from libraries
5. Deliver a timely service for digital copies within short time (few weeks)
6. Digital copy cost should not be more than a print copy, if available, max £100
7. On demand service (eg digitisation/provision of digital copy) seemed most appropriate route for ad hoc requests, through existing mechanisms eg BL
8. Appropriate format: searchable pdf as minimum but epub or HTML5 preferred. Accessibility for users with disabilities a big problem, eg screen readers etc
Mapping publicly available digitised books

- Related project evaluating sources of digitised openly available books with the aim to link NBK records to existing digitised copies
- Starting with major sources (Hathi, IA, Project Gutenberg) and moving on to others (e.g., Europeana, DPLA...)
- Key challenge in identifying books: different terminology used to express licences (PD, CC, Not in copyright...)
- Other issues to consider will be:
  - Overlap/de-duplication?
  - Trust and persistent URLs
  - APIs
  - Of course part of a solution already exists through WorldCat>Hathi Trust relationship
Next steps

Digital access pilots
• final report and analysis of data will provide us with a good indication of the state and size of “the problem”
• phase 2: approach top ten publishers to explore the issues and see what the response is. Then plan next phase accordingly.

NBK
• carry out dialogue with potential suppliers over the Summer, formal procurement in progress, supplier selection in Oct 2016
• clarify relationship between NBK and increasing access to digital books (e.g. availability data, licensing, workflow...)
Image credits

Coin by Kirill Tomilov from the Noun Project CC
https://thenounproject.com/search/?q=coin+two+sides&i=160453

Help (question mark) By Consumer Financial Protection Bureau, US from the Noun Project Public domain
https://thenounproject.com/search/?q=question+mark&i=89611

Work in progress by Gleb Khorunzhiy from the Noun Project CC
https://thenounproject.com/search/?q=works+in+progress&i=215094

Alert by Alain LOUBET from the Noun Project CC https://thenounproject.com/search/?q=health+warning&i=14055
Thank you

Any questions?

Neil Grindley, neil.grindley@jisc.ac.uk @neilgrindley
Paola Marchionni, paola.Marchionni@jisc.ac.uk @paolamarchionni