Share the Experience: Collection Management
CM@Hull

7th September 2017

The National Bibliographic Knowledgebase

Neil Grindley, Jisc Head of Resource Discovery
Bethan Ruddock, Jisc NBK Project Manager
Investing in library services

National bibliographic knowledgebase

We’re building a new UK-wide service to help libraries collaboratively manage their collections and improve access to print and digital resources.

We’re using the latest digital technologies to create a national bibliographic knowledgebase.

Transforming our library support services

Developing a more effective, efficient and cohesive set of library services.

Started 1 May 2016
Budget: £250k to £1m
Expected outcome: Service development

Tags: DATA AND ANALYTICS, LIBRARIES, RESOURCE DISCOVERY
What is the National Bibliographic Knowledgebase?

- Legal Deposit Libraries
- Academic Research Libraries
- Teaching & Business Engaged Libraries
- Specialist Libraries
- Open Access Book Directories
- Availability Data
- Usage Data
- Benchmarking Collections
- Managing Print Collections
- Document Delivery
- Digitisation and Preservation
- Resource Discovery
- Copy Cataloguing
What is it?

The National Bibliographic Knowledgebase (NBK) will aggregate bibliographic data at scale and link with a number of other data sources to inform library collection management decisions and to help users more effectively find, access and use print and digital scholarly resources.
What is the National Bibliographic Knowledgebase?

Availability Data

Usage Data

Legal Deposit Libraries

Academic Research Libraries

Teaching & Business Engaged Libraries

Specialist Libraries

Open Access Book Directories

Benchmarking Collections

Managing Print Collections

Document Delivery

Digitisation and Preservation

Resource Discovery

Copy Cataloguing

Open Access Book Directories
What are we trying to achieve?

The NBK will help transform how libraries manage their collections, provide access to resources and collaborate with each other. It will provide a sustainable fit-for-purpose next generation national data infrastructure that practically supports libraries to make the transition from a print-first to a digital-first paradigm.
Academics, Researchers & Learners

Publishers & Data Suppliers

Academic & Specialist Libraries

NBK

CBS System

Infrastructure

Synchronisation

WorldCat®
NBK is an engagement programme
... and a sustainability challenge

Open Services

Login required

Chargeable Services
<table>
<thead>
<tr>
<th>Open Free Services</th>
<th>Login Required</th>
<th>Chargeable Services</th>
</tr>
</thead>
<tbody>
<tr>
<td>Function</td>
<td>Purpose</td>
<td>Function</td>
</tr>
<tr>
<td>Data Ingest</td>
<td>Upload Library catalogues</td>
<td>International collection benchmarking tools</td>
</tr>
<tr>
<td>Discovery</td>
<td>Find scholarly materials</td>
<td>Bulk data download</td>
</tr>
<tr>
<td>Basic cataloguing tool</td>
<td>Create catalogue records</td>
<td>Data clean up &amp; workflow consultancy</td>
</tr>
<tr>
<td>General user support</td>
<td>Navigate between services</td>
<td>Collection development consultancy</td>
</tr>
<tr>
<td>NBK blog &amp; info pages</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jisc x-service dashboard</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Contributor dashboard</td>
<td>Check contributor status</td>
<td></td>
</tr>
<tr>
<td>Data troubleshooting</td>
<td>Improve workflow</td>
<td></td>
</tr>
<tr>
<td>Catalogue record download</td>
<td>Enhance catalogue</td>
<td></td>
</tr>
<tr>
<td>Catalogue record upload</td>
<td>Dynamically update aggregation</td>
<td></td>
</tr>
<tr>
<td>Collection management tools</td>
<td>Manage collections</td>
<td></td>
</tr>
<tr>
<td>NBK usage data</td>
<td>Check and compare value of resources</td>
<td></td>
</tr>
<tr>
<td>Researcher tools</td>
<td>Exploit bibliographic data</td>
<td></td>
</tr>
</tbody>
</table>

NBK Participant Framework 1
Collections Management (or not)

Collection Assessment
- Increased discoverability of, and opportunity to showcase, special collections.
- Increased discoverability of all library resources through Jisc NBK Discovery, WorldCat, and search engines.
- Reports about data quality allowing libraries to focus their efforts on data improvement where needed
- Potential to have ‘enhanced’ data pushed back to libraries
- Possibility of aggregated usage data to inform purchasing decisions
- Potential to link with JUSP and KB+ to help manage e-resources
- Allow university presses to contribute their data, widening audience and impact
- Potential for consortial interfaces, without libraries needing to procure & manage a shared catalogue
- Will include OA material, allowing libraries to more easily see if there is an OA version of a work before purchasing

Collection Benchmarking

Collection Comparison
Understanding collections overlap: an investigation into White Rose Libraries collections using the SCS GreenGlass and COPAC Collaboration Collection Management Tool

Final Report, July 2017

Background 2
Understanding record matching in GreenGlass 3
Understanding record matching in the Copac CCM Tool 5
Summary of Data checking undertaken 5
Test 1 5
Background

In 2016, WRL began work with SCS GreenGlass to analyse their collections and explore collaborative collection management between the 3 libraries. The collection overlap reported by GreenGlass was considerably lower than anticipated.

- 83% of the titles in the 3 libraries’ combined collections were identified as uniquely held by only 1 library (1,638,570 of 1,971,001 titles held)
- 75.5% of University of Leeds collections identified as unique within WRL
- 64.6% of University of Sheffield collections identified as unique within WRL
- 58.9% of University of York collections identified as unique within WRL
Methodology

Supported by Jisc funding, WRL did a ‘deep dive’ into a selection of data to compare results of collection overlap between:

• GreenGlass
• Copac Collection Management Tool (CCM)
• A manual examination of records
“It is worth stating early on in this report that our data checking exercises have shown that GreenGlass has actually given us a reasonably accurate picture of the overlap between the WRL collections. It is also worth noting that the overlap between WRL is similar to that of US libraries who have already used the GreenGlass tool.

We did identify that our overlap was slightly underreported, and established that the reasons for that underreporting were largely due to incomplete and inconsistent metadata in the WRL catalogue records.”
The Physics results indicated:

• The presence or absence of an ISBN in the record has minimal impact (1-2%) on matching accuracy

• Manual Excel checking closely reflects the GreenGlass results, showing only a 2-4% difference from the GreenGlass totals

• The CCM Tool results differ by 11-12% difference from the GreenGlass totals. (We investigated this, and found that the discrepancy was due to records for items in York’s External Store had not been exported to Copac)

• Overall therefore, the GreenGlass matching seems to differ by only small percentages from other methods tested

• However, small percentages translate into large numbers of books.
3 main categories of factors affecting the matching of White Rose Libraries Records

1. **Data preparedness and exports**: the profiles for regular exports to external catalogues (Copac, WorldCat, etc.) need to be understood and re-assessed when a new project is undertaken, rather than simply copied across. The amount of uncatalogued material should also be understood and highlighted.

2. **Metadata quality**: not only does current cataloguing practice vary between UK libraries, each library will also have records reflecting a variety of different legacy approaches to cataloguing.

3. **Matching algorithms**: it is worth taking time to investigate how matching works in different tools, and to encourage the providers of those tools to experiment with different algorithms.
Recommendations

1. Share a version of this report more widely
2. Develop workflow guidance and best practice to help libraries export data to external catalogues
3. External catalogues should indicate the completeness and currency of contributing library holdings
4. Develop guidance to help libraries understand the impact that metadata quality has on matching records
5. Investigate ways to help libraries improve the quality of their metadata in catalogue records
6. Develop advice and guidance for libraries using collection analysis tools
7. Develop advice and guidance for libraries embarking on collaborative collection management initiatives
8. Libraries to contribute to the future development of collection analysis tools
Timeline

Jan 2017
Project initiation

Jun 2017 - Jan 2018
Receive and load phase 2 datasets

2018
Receive and load phase 3 datasets.
Interface development.

2019
Receive and load phase 4 datasets.
Enhancement work.

Jan 2018
NBK beta interface launch.
75 libraries on NBK

July 2018
110 libraries on NBK

Jan 2019
NBK is live. Copac, RLUK database and SUNCAT replaced by NBK.
150 libraries on NBK

Jan 2020
Project end.
NBK is in BAU.
225 libraries on NBK.

July 2017
Trial datasets loaded.
10 libraries on NBK
How does data get into the NBK?

Libraries ➔ Send data to Jisc ➔ Process data ➔ Load data onto NBK

- Send data to Jisc
- Minimal pre-processing, as required
- Send data to
- We will convert non-MARC data to MARC format
- Process data
- Load data onto

All the data that’s visible on your local catalogue. We can suppress records if you can’t drop them. Include holdings information, journals, ejournals, ebooks
How does data get out of the NBK?

- Libraries: Send data to Jisc
  - Jisc: Minimal pre-processing, as required
  - Send data to OCLC
  - OCLC: Process data
  - Load data onto NBK
  - NBK: Load data onto WorldCat
  - WorldCat: Jisc resource discovery service
  - Selective sync to Other aggregators
  - Jisc cataloguing service

Other aggregators
What information do libraries get about data?

Libraries → Send data to Jisc → Minimal pre-processing, as required → Send data to OCLC → Process data, Load data onto NBK → Other aggregators → Selective sync to Jisc cataloguing service → WorldCat → Error reports
Collection Management & other services

- Libraries
  - Send data to

- Jisc
  - Minimal pre-processing, as required
  - Send data to

- OCLC
  - Process data
  - Load data onto

- NBK
  - Jisc cataloguing service
  - Other services?

- CCM Tools
  - Freely available

- Jisc resource discovery service
  - Free for Jisc members

- Jisc resource discovery service
  - Other services?
Collection Management

- Send data to Libraries
- Send data to Jisc
  - Minimal pre-processing, as required
  - Send data to OCLC
- Process data
- Load data onto NBK

Until ~2019

- Copac
- CCM Tools
- Jisc cataloguing service
- Other services?

Login (free for Jisc members)
The NBK will be a larger bibliographic database than the current Copac database. This means more libraries as sources/benchmarks/comparators/collaborators for collection management.

Most data will be being added over the next 2.5 years. CCM Tools will continue to use the Copac database for now. We will start testing CCM Tools with the NBK in late 2017.
<table>
<thead>
<tr>
<th>NBK libraries so far</th>
<th>Upcoming NBK libraries</th>
</tr>
</thead>
<tbody>
<tr>
<td>British Library</td>
<td>Leeds</td>
</tr>
<tr>
<td>Oxford</td>
<td>QUB</td>
</tr>
<tr>
<td>National Trust</td>
<td>Cardiff Uni</td>
</tr>
<tr>
<td>York</td>
<td>Aberystwyth</td>
</tr>
<tr>
<td>Leicester</td>
<td>RWCMD</td>
</tr>
<tr>
<td>Manchester</td>
<td>Aberdeen</td>
</tr>
<tr>
<td>TCD</td>
<td>Bristol</td>
</tr>
<tr>
<td>Sheffield</td>
<td>Royal College of Music</td>
</tr>
<tr>
<td>Senate House</td>
<td>The National Archives library</td>
</tr>
<tr>
<td>Glasgow</td>
<td>UWTSD</td>
</tr>
<tr>
<td></td>
<td>St Andrews</td>
</tr>
<tr>
<td></td>
<td>London Met</td>
</tr>
<tr>
<td></td>
<td>Bangor</td>
</tr>
<tr>
<td></td>
<td>Glyndwr</td>
</tr>
<tr>
<td></td>
<td>Cardiff Met</td>
</tr>
<tr>
<td></td>
<td>NLW</td>
</tr>
<tr>
<td></td>
<td>Swansea</td>
</tr>
<tr>
<td></td>
<td>Royal Holloway</td>
</tr>
<tr>
<td></td>
<td>Southampton</td>
</tr>
</tbody>
</table>
Get involved

» Library not on those lists?
  › Contact nbk.copac@jisc.ac.uk and we’ll start the process

» Other involvement
  › We’ll be looking for community feedback and involvement on the NBK & its services, including collection management
  › Contact nbk.copac@jisc.ac.uk if you’d like to be involved