Not just a British library: enabling a global discovery experience

Dimity Flanagan
Scholarly Communications Lead
0000-0002-6784-0325 / @2warpthreads
‘Open’ and the British Library

• In 1972, the British Library Act established the British Library as “… a national centre for reference, study and bibliographical and other information services” and tasks us to contribute “to the efficient management of other libraries and information services, to carry out and sponsor research”.

• In 2015, the Living Knowledge vision stated that the best way to deliver this mission today is for the British Library to become the most open, creative and innovative institution of its kind in the world.
Repositioning the BL in the research landscape

“It must be the most reliable in terms of finding relevant information. It has a lot of catalogues to search for information. It has the most dedicated staff to meet demands on the national and international level. It has excellent facilities and gives both physical and online access.”

“Old-fashioned.”

“The British Library sounds like an impressive building with an extensive collection of resources, somewhere I would like to visit but I am not based in London.”

“Very large institute with every book ever published.”
Changing nature of collections

- Relative to content on the internet our collections are shrinking
- We can no longer hope to meet all user needs from our collections
- Enhance our collections by connecting them to relevant global content
- We need to reconsider national collection models in a global open science world
Improving open access discovery

Explore the British Library

Open access
Legal Deposit digital
Subscribed digital
Physical Reading Room Access
BL On Demand
It’s not open if you cannot find it...

Open Access Discovery Roadmap 2018

<table>
<thead>
<tr>
<th>Project</th>
<th>Organisation</th>
<th>Timescale</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rebuild and enhance the BASE server infrastructure</td>
<td>BASE</td>
<td>2018</td>
</tr>
<tr>
<td>Prepare BASE index structure for the next ORCID.DE project</td>
<td>BASE</td>
<td>2018-2019</td>
</tr>
<tr>
<td>Build a repository/data source registry or directory for open access content</td>
<td>BASE</td>
<td>2019</td>
</tr>
<tr>
<td>Open access discovery integration with BL search systems</td>
<td>British Library</td>
<td>2019-2020</td>
</tr>
<tr>
<td>Next generation pilot projects for repositories - improving repository software platforms to support improved discovery</td>
<td>COAR</td>
<td>2018-2019</td>
</tr>
<tr>
<td>International adoption of common metadata standards to support tracking OA</td>
<td>COAR</td>
<td>2018-2020</td>
</tr>
<tr>
<td>Aligning regional repository networks (aggregators) to support metadata exchange and increased visibility of repository resources</td>
<td>COAR</td>
<td>2018-2020</td>
</tr>
<tr>
<td>Integration of CORE data with open citation datasets in collaboration with Microsoft</td>
<td>COBI</td>
<td>2018-2019</td>
</tr>
<tr>
<td>Indexing all documents listed in Crossref, for which we can locate a freely available copy anywhere on the Internet</td>
<td>COBI</td>
<td>2019</td>
</tr>
<tr>
<td>Working with publishers to increase the quality and proliferation of licenc metadata they register</td>
<td>Crossref (and others)</td>
<td>2018-2020</td>
</tr>
<tr>
<td>Improved metric reports</td>
<td>DOAJ</td>
<td>2020</td>
</tr>
<tr>
<td>Upgrade of Elasticsearch to improve search and UI</td>
<td>DOAJ</td>
<td>2020</td>
</tr>
<tr>
<td>New manuscript submission system</td>
<td>Europe PMC</td>
<td>2019</td>
</tr>
<tr>
<td>Ingest of additional preprint metadata and full text of preprints</td>
<td>Europe PMC</td>
<td>2019</td>
</tr>
<tr>
<td>Submission system for text-mining groups to submit annotations of articles in Europe PMC</td>
<td>Europe PMC</td>
<td>2018 (complete)</td>
</tr>
<tr>
<td>Get the Research (scholarly search engine for the general public, built on Unpaywall index)</td>
<td>Impactstory</td>
<td>2018-2019</td>
</tr>
<tr>
<td>Unpaywall to include papers without DOIs</td>
<td>Impactstory</td>
<td>2019</td>
</tr>
<tr>
<td>Building a research software citation index</td>
<td>Impactstory</td>
<td>2019</td>
</tr>
<tr>
<td>Measurement dashboards to encourage OA mandate compliance</td>
<td>Impactstory</td>
<td>2018-2019</td>
</tr>
<tr>
<td>Open authorship/citation dataset (disambiguated author clusters, citations, institutions)</td>
<td>Impactstory</td>
<td>2019</td>
</tr>
<tr>
<td>FatCat scholarly record preservation and metadata aggregation project</td>
<td>Internet Archive</td>
<td>2018-2020</td>
</tr>
<tr>
<td>At-Risk OA journals preservation/access project</td>
<td>Internet Archive</td>
<td>2018-2020</td>
</tr>
<tr>
<td>Archiving and indexing services for OA discovery</td>
<td>Internet Archive</td>
<td>2018-2019</td>
</tr>
<tr>
<td>National Bibliographic Knowledgebase- UK-wide service to help libraries collaboratively manage their collections</td>
<td>Jisc</td>
<td>2018-2019</td>
</tr>
<tr>
<td>Working with stakeholders in the library supply chain to support the discovery of OA monographs, particularly for New University Presses (NUPs) and Academic-Led presses</td>
<td>Jisc</td>
<td>2018-2019</td>
</tr>
<tr>
<td>Dynamic Purchasing System call for OA publishing services via the OIEU</td>
<td>Jisc</td>
<td>2018-2019</td>
</tr>
<tr>
<td>Toolkit for New Academic Presses, which will include model licenses and best practice</td>
<td>Jisc</td>
<td>2018-2019</td>
</tr>
<tr>
<td>Jisc is working with ORCID, Freya project, ARDC and others to pursue how PIDs can be better used to address the discovery challenge</td>
<td>Jisc</td>
<td>2018-2019</td>
</tr>
<tr>
<td>GePDF. Libraries can provide PDF access to their patrons anytime, anywhere, in a click with or without subscriptions</td>
<td>Open Access Button</td>
<td>2018-2019</td>
</tr>
<tr>
<td>InstantTTL: instantly fills forms and pre-checks against subscriptions and OA</td>
<td>Open Access Button</td>
<td>2019</td>
</tr>
<tr>
<td>Improved UI for deposit into repositories</td>
<td>Open Access Button</td>
<td>2019</td>
</tr>
</tbody>
</table>
Partnership example: Impactstory
Single Digital Presence for public libraries?

Options  User needs  Strategic Approach  Options Analysis  Feasibility Testing

Developing an open repository platform

- Single BL repository platform
- Refresh national preservation system (>5m items, petabyte-scale)
- Access layer with multiple repositories, shared service model
- Repository pilot developed with:

![Diagram](image)

- Access Layer
- Services Layer
- Preservation Layer

Repositories involved:
- ETHOS
- Data.bl.uk
- BL Institutional Repository
- Partner Repositories
Project FREYA

- FREYA is a 3-year project funded by the European Commission under the Horizon 2020 programme. The project aims to extend the infrastructure for persistent identifiers (PIDs) as a core component of open research, in the EU and globally. FREYA will improve discovery, navigation, retrieval, and access to research resources.
# Enhancing metadata: provenance

<table>
<thead>
<tr>
<th>Alternate identifier</th>
<th>identifier: DAR02144</th>
<th>type: Digital Asset Register ID</th>
</tr>
</thead>
</table>
| Related identifier    | identifier: http://access.bl.uk/ark:/10850/da_100022500001 0x000002 | type: ARK  
relation: HasPart |
|                       | identifier: http://access.bl.uk/ark:/10850/da_100022500002 0x000002 | type: ARK  
relation: HasPart |
|                       | identifier: http://access.bl.uk/ark:/10850/da_100022500003 0x000002 | type: ARK  
relation: HasPart |
|                       | identifier: http://access.bl.uk/ark:/10850/da_100022500004 0x000002 | type: ARK  
relation: HasPart |
|                       | identifier: http://access.bl.uk/ark:/10850/da_100022500005 0x000002 | type: ARK  
relation: HasPart |
|                       | identifier: http://access.bl.uk/ark:/10850/da_100022500006 0x000002 | type: ARK  
relation: HasPart |
|                       | identifier: http://access.bl.uk/ark:/10850/da_100022500007 0x000002 | type: ARK  
relation: HasPart |
|                       | identifier: http://access.bl.uk/ark:/10850/da_100022500008 0x000002 | type: ARK  
relation: HasPart |
|                       | identifier: http://access.bl.uk/ark:/10850/da_100022500009 0x000002 | type: ARK  
relation: HasPart |
|                       | identifier: http://access.bl.uk/ark:/10850/da_100022500010 0x000002 | type: ARK  
relation: HasPart |
|                       | identifier: http://access.bl.uk/ark:/10850/da_100022500011 0x000002 | type: ARK  
relation: HasPart |
|                       | identifier: http://access.bl.uk/ark:/10850/da_100022500012 0x000002 | type: ARK  
relation: HasPart |
|                       | identifier: http://access.bl.uk/ark:/10850/da_100022500013 0x000002 | type: ARK  
relation: HasPart |
|                       | identifier: http://access.bl.uk/ark:/10850/da_100022500014 0x000002 | type: ARK  
relation: HasPart |
Cite the digital

- Our URLs for digital content can be long
- Our ARKs are not yet resolvable
- There is a tendency to cite print versions rather than digital because:
  - Concerns over persistency of URLs
  - URLs are too long
  - Print is seen as more reliable


Technical and cultural changes needed
Text and Data Mining: Newspapers

- Much content can only be accessed in BL Reading Rooms
- OCR quality – substantial digitisation done with Findmypast was from microfilm which does impact on accuracy. Furthermore, we do not have access to the derived data (OCR).
- Transparency over selection decisions (how to convey in final dataset?)
- Derived data from the digitisation (OCR, entities etc) will be made freely available via the http://data.bl.uk platform.
Dimity Flanagan
Scholarly Communications Lead
0000-0002-6784-0325 / @2warpthreads