Making The Connection

Creating a Linked Open National Bibliography

Neil Wilson
Head, Metadata Services

DPC Open Data Briefing Day
July 5th 2013
The British Library…

- Receives a copy of every publication produced in UK

- Collection includes over 150 million items:
  - 14 million books
  - 920,000 journal & newspaper titles
  - 58 million patents
  - 3 million sound recordings
  - …and much more

- Operates the world's largest document delivery service
Metadata Provision
BL Responsibilities

• The BL Act records our role “as a national centre for reference, study and bibliographical and other information services…”

• BL Metadata Services support these core activities by:
  • Enabling provision of national & international bibliographic services
  • Increasing visibility of holdings to connect users to content
  • Enabling participation in international collaborative initiatives
A Changing Environment
Library Sector Relevance

Declining?

“I did my PhD with only 12 visits to a library. That was 5 years ago; things have improved since then, now you don’t need to use a library at all!”

Increasing?

“The release of library data offers the opportunity for it to be used in ways unthought of by the library & information community…”

Opening library data

Libraries generate and hold large amounts of data. David Stuart looks at the potential benefits of sharing and allowing people to innovate with it.

Research Information: April/May 2011

An increasingly wide range of organisations are starting to recognise the advantages of making some of their data publicly available online, and allowing external individuals and organisations to innovate around it. Although such practices were once primarily associated with leading-edge web organisations such as Google and Amazon, it is increasingly expected to be part of every organisation’s online strategy. This ranges from supermarkets such as Tesco (www.tesco.com/forum) looking to exploit the commercial potential of making data available, to government departments responding to demands for increased transparency and access to potentially valuable publicly-funded data (for example, data.gov.uk).
A Changing Environment

Our users were once lone scholars but now include:

• Collaborative Learners
• Data miners
• Machines

Libraries must respond

• Improving relevance
• Stimulating innovation
• Enabling knowledge creation
A Changing Environment
UK Public Sector Open Data

- The Web - fostered expectation information should be open to all

- Data has new value - from open technologies enabling rapid connection & reuse

- ‘Open Data’ - has evolved to cover:
  - Transparency
  - Economic growth
  - Research & discovery
A Changing Environment

UK Public Sector Open Data Rating

UK Government commitment to the principle of open public data has grown since 2009 with a proposed rating:

* **Available on Web** + open licence

** ** + machine-readable

*** + non-proprietary format

**** + W3C open standards

***** + link to other data
Open Metadata Strategy
Risk Management

- **Financial** - *Loss of revenue from legacy services*

- **Legal** - *Many licensing options exist but few organisations offer unconditional universal access*

- **Reputational** – *Perception we don’t meet government & community expectations e.g. via differing interpretations of ‘free’, ‘open’ & ‘linked’*

<table>
<thead>
<tr>
<th></th>
<th>Free</th>
<th>Open</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Of cost?</td>
<td>To access from anywhere</td>
</tr>
<tr>
<td></td>
<td>Of all licensing restrictions?</td>
<td>To modification/enhancement</td>
</tr>
<tr>
<td></td>
<td>To sell or re-use in commercial services</td>
<td>To use with free software</td>
</tr>
<tr>
<td></td>
<td>To copy &amp; redistribute?</td>
<td>To link to/from your data or system</td>
</tr>
</tbody>
</table>
Linked / Open Data?

- **Open Data** – can be made available *without* being linked

- **Linked Data** – can be created *without* being open

- **Linked Open Data** – *Is open data created using techniques enabling it to be linked with other datasets*

- **Both** – may be available as file dumps and/or live services

*See: [http://www.semantic-web.at/LOD-TheEssentials.pdf](http://www.semantic-web.at/LOD-TheEssentials.pdf)*
The BNB & Linked Data

Why BNB?

We wanted to:

• **Advance debate** from theory to practice via release of a ‘critical mass’ of data

• **Show commitment** by using a large, core dataset; niche examples are not as compelling

• **Create a foundational service** others can build on & not a dead end

• **Offer a reusable dataset** of published items & not a unique institutional catalogue
BL Open Metadata Strategy
Meeting the Challenge

Service Aims:

- **Collaboration** - on innovative new services advancing debate from theory to practice

- **Remove barriers** - enabling innovation & reuse without unnecessary restrictions

- **Recognition** - while adopting open licensing (CC0) appropriate to the widest re-use
Why Linked Data?

See:
http://vimeo.com/36752317
Linked Open Data Cloud

September 2011
Linked Data
How is it different?

• Traditional library metadata is based on a static, proprietary **document based** model (e.g. catalogue card)

• Linked Data uses a dynamic **data based** model based on linking simple statements about resource properties (*Triples*)

<table>
<thead>
<tr>
<th>Traditional Metadata</th>
<th>Linked Metadata</th>
</tr>
</thead>
<tbody>
<tr>
<td>Passive</td>
<td>Dynamic / Interactive</td>
</tr>
<tr>
<td>Self-contained</td>
<td>Linked to External Resources</td>
</tr>
<tr>
<td>End Result</td>
<td>Options for Further Inquiry</td>
</tr>
<tr>
<td>Domain Specific Standards</td>
<td>Open Structure &amp; Standards</td>
</tr>
</tbody>
</table>
A Change in Emphasis

• From self contained records describing resources

• To simple statements about resource properties (e.g. [This book] [has the author] [Charles Dickens])

• With ‘records’ assembled from selected statements
Remodelling The Data
Simple Triple Statements

- Title: Great Expectations
- Author: Dickens, Charles 1812-1870
- Subject: England - Social conditions - Fiction
- Publisher: Penguin English Library
- ISBN: 9780141198897
- Pagination: 570p
Common Vocabularies
Selection

- Bibliographic Ontology
- Bio: a Vocabulary for Biographical Information
- British Library Terms
- Dublin Core
- Event Ontology
- FOAF: Friend of a Friend
- ISBD
- Org: an Organisation Ontology
- OWL
- SKOS
- RDF Schema
- WGS84 Geo Positioning
Linking

To give our data broader context we linked to:

• General resources:
  • GeoNames
  • Lexvo
  • RDF Book Mashup

• Library resources:
  • LCSH
  • VIAF
  • Dewey.info
The BNB & Linked Data
Remodelling from MARC to RDF
Linked Data
Potential Benefits

**Advantages:**
- Multilingual Approaches
- Personalisation
- Inferencing
- Disambiguation

**Issues:**
- Trust
- Persistence
- A work in progress…

But don’t confuse the technology with the ideology…
Where Did We Get To?

Multiple Access Routes

- thedatahub.org/dataset/bluk-bnb-basic
- thedatahub.org/dataset/bluk-bnb

British National Bibliography (BNB) - Basic Representation

Resources
- bnb.data.bl.uk/sparql
- bnb.data.bl.uk

BNB 1950-2012
2.8 Million Records
85 Million Unique Triples

Welcome to bnb.data.bl.uk

The BNB Linked Data Platform provides access to the British National Bibliography published as linked open data and made available through SPARQL, services. Two different interfaces are provided: a SPARQL editor, and /sparql a service endpoint for remote queries. Alternatively, use the search box below to enter a plan text term.

The British National Bibliography (BNB) records the publishing activity of the United Kingdom and the Republic of Ireland and has been doing so since the 1950s. This has traditionally included printed works and has recently been extended to electronic publications. The dataset includes metadata about published books, already published and forthcoming, and serials i.e. journals, periodicals, magazines, newspapers, etc.

What is available?

- The Linked Open BBN is a subset of the full British National Bibliography. It currently includes published books (including monographs published over time) and serial publications, representing approximately 2.6 million records.

The dataset is available under a Creative Commons
BL Open Metadata Strategy
Achievements

• 900 organisations in over 90 countries use free services e.g. Open Library, BBC & Wikimedia Commons

• Worked with Government, W3C & developers on technical, standards & licensing issues

• 95% satisfaction rate - from 2012 survey

• Creation of Linked Open BNB – cited by Cabinet Office as best practice example
BL Linked Open Metadata

Benefits

• **Presence, visibility** & ‘first mover advantage’

• **BL data models** – influential on wider developments

• **New collaborations** - with public & private sector groups e.g. Microsoft

• **New staff expertise**

• **Confirmation that valuable data will be used** – with up to 9 million BNB transactions monthly
Linked Open Data Challenges

How to?

• **Capture value & gain attribution for effort despite ‘cloning’!**

• **Ensure expectations are met - Government & wider community**

• **License 3rd party data - to support open re-supply**

• **Ensure effective service maintenance - & persistence**
Conclusions
Linked Open Data

• Open Data offers knowledge organisations new opportunities - authority, persistence & stability being seen as invaluable for future information infrastructure

• Lower barriers for data re-use increase its community value - & improves access to information/culture

• Improved web integration of resources increases visibility - & reaches new users

• A global pool of reusable data – creates new opportunities & enables organisations to add unique value
Thank You

Questions?

http://twitter.com/#!/BLMetadata
metadata@bl.uk
http://www.bl.uk/bibliographic/datafree.html