Sharing is good: getting the most out of your library’s data

Bethan Ruddock  @bethanar

https://www.flickr.com/photos/ryanr/142455033
Data Sharing

Where do you share your data?

Personal data
Consumer data

Is it a choice?
What do you get in return?
What do we want to achieve?
How can sharing data help us achieve this?
### Passive data sharing vs Active data sharing

<table>
<thead>
<tr>
<th>Passive data sharing</th>
<th>Active data sharing</th>
</tr>
</thead>
<tbody>
<tr>
<td>No specific goal/use</td>
<td>Specific goal/use</td>
</tr>
<tr>
<td>Speculative, ‘see what happens’</td>
<td>Know what you want to happen</td>
</tr>
<tr>
<td>Share with everyone</td>
<td>Share with specific partners</td>
</tr>
<tr>
<td>Open format (eg API, triple store)</td>
<td>Appropriate formats</td>
</tr>
<tr>
<td>Open data licence</td>
<td>Appropriate data licence</td>
</tr>
<tr>
<td><a href="http://opendatacommons.org/licenses/odbl/">http://opendatacommons.org/licenses/odbl/</a></td>
<td><em>(maybe) less effort</em></td>
</tr>
<tr>
<td><em>(maybe) less effort</em></td>
<td><em>(maybe) more effort</em></td>
</tr>
<tr>
<td>Less curation</td>
<td>More curation</td>
</tr>
</tbody>
</table>
Why share data?

“The best thing to do with your data will be thought of by someone else”  Rufus Pollock

“The best thing to do with your data will happen when it meets someone else’s data”
How can you get your data out?
» Can the data be exported?

» What export formats are available?
  › Are these standard formats?
  › Are they transformable and reuseable?

» How does the exported data need to be stored?

» Can you export the data yourself?
  › If not, is there a charge for export?
Ask yourself

» Do you know what to do with the exported data?

» Do you have the expertise to work with it or transform it?
  › What additional expertise might you need? (eg programming, statistics)

» Do you know what questions to ask of the data?
  › Do you know how to understand/interpret the answers?
What can you do with shared library data?
Physical union catalogue
Data comes in in MARC, XML - & others!
Data must be consistent, transformable
Data is matched on certain fields, to create ‘deduplicated’ merged records.
“a link to Higher Education and other research libraries, and that’s good because [we] are isolated in many ways. We mention it to visitors – gets us a bit of kudos!” (Specialist library)
“Making our resources identifiable to new audiences and other professionals.”
(Specialist library)
“Allows [us] to identify possible partners for collaboration.” (Specialist library)
“Very useful for academics and post-graduates – and feeling of supporting the scholarly community.” (University library)
We can’t really justify ourselves to the [governing body] because they want a different readership, or usership – I mean they’re after tourists specifically. We’ll never get thousands of tourists or visitors through our door, because we’ve got specialised book collections. So we’ve got to reach out ‘over the hills and far away’ which is what you [Copac] help us do." (Specialist library)
"We’re not the most approachable library; [specialist area] libraries aren’t, by the very nature of things, so this is another **branch out into the wider world**, which is so important for us…Copac reaches that wider audience. Cataloguing is important to me because otherwise the books just gather dust. I don’t want that: **I want them used and read.** (Specialist library)
“Prestige. Being part of the research community and having what you know are very valuable holdings exposed to the research community …. Potential for collaboration in the future – in collection management? … Envisage eventually being able to compare holdings using a central tool.” (University library)
Collaborative collections management:
• Identify last copies.
• Assess material for retention, digitisation, and conservation.
• Visualise collection strengths.
• Discover rare and unique holdings.
• Develop new collections.
• Explore subject collections across the UK
What: Copac Collections Management Tools (CCM)
“The tools’ *ease of use* has been impressive….we can already see the benefits these tools can provide: a *better understanding of our collections* in comparison with other HE institutions; the ability to *highlight and promote research strengths* to our user community, and support for our *collection management decisions*. ”
"Overall, the actual use of the Tool took just over an hour of staff time, for each of the trials. Without the Tool, and to use similar withdrawal criteria as above, the online Copac database would have to have been searched for each of the 6,836 items initially submitted. Based on previous experience of an average of 3 minutes a search, this would have taken about 342 hours just to do the search."
**What:** the Archives Hub

**Archive descriptions in EAD, an XML format**

**Descriptions exported in or created in EAD**
1) Processing scripts

Scripts can only be applied in certain circumstances, where we can see patterns that allow use of a transformation script, and the required outputs must be consistent.
2) Manual edits
Where the change that we require is in the content of the description rather than the structure of the EAD. This is usually used for descriptions that do not meet our requirements for mandatory fields.
3) No action
Where discrepancies do not affect the validity of the EAD document, the Archives Hub functionality, or the usability of the finding aid for Archives Hub users, we do not take any action to change them.
Why contribute to the Archives Hub?

"[Contributing to the Archives Hub] increases the number of collections catalogued online, leading to more users, leading to more impressive stats for fundraising. On this basis, Archives Hub is highly important in the strategic development of our service." (Museum archive)
“[The Archives Hub’s] existence means for the first time we can have an archival catalogue that exposes our collections, used in conjunction with the digital archive on the repository, means for the first time internationally important collections will be discoverable and accessible.” (Higher Education archive)
"Good to be able to **show it to management/stakeholders** as proof of work done. It **looks professional** and the **exciting projects** done by Hub (like sharing data with Archives Portal Europe) sound good to management ears. And it's free." (Higher Education archive)
"I think it's good for our collections to be presented in context of other UK collections." (Higher Education archive)
Why contribute to the Archives Hub?

"I think people who have not used archives before may be nervous and put off because they don't have the chance to familiarise themselves with the catalogue and plan their visit in advance. Being able to browse archive catalogues online is akin to going onto a shop website before entering the intimidating showroom!" (Museum archive)
LAMP/Learning analytics: use cases

» Merge data from multiple systems: library, student registry, IT services
» Contribute to the institutional analytics mission
» Avoid data and reporting silos i.e., data sets, spreadsheets and reports
» Compelling visualisations
» Map e-resource usage to actual users
» Key usage indicators by discipline
» Identify usage levels i.e., high or low
» Analyse library gate data
» Examine events by specific user groupings i.e., overseas, distance
Collect comprehensive and granular eresource data

Merge data from the student registry

Identify how different faculties use the library and the collection

Identify least-used resources/where usage is less/more than expected

Produce compelling visualisations

Provide a dashboard with key indicators by department

Correlate feedback (e.g. NSS), enquiries and collection strength

Correlate reading lists with actual usage

Examine events by specific demographics/groupings
<table>
<thead>
<tr>
<th>Mode of Attendance</th>
<th>Description</th>
<th>Attendance Mode</th>
<th>Full-time</th>
<th>Part-time</th>
<th>Sandwich</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 Attendance mode</td>
<td>H</td>
<td>standard</td>
<td>full time</td>
<td>full time</td>
<td></td>
</tr>
<tr>
<td>2 Attendance mode</td>
<td>H OT</td>
<td>standard</td>
<td>full time</td>
<td>full time</td>
<td></td>
</tr>
<tr>
<td>2 Attendance mode</td>
<td>H PP</td>
<td>standard</td>
<td>full time</td>
<td>full time</td>
<td></td>
</tr>
<tr>
<td>2 Attendance mode</td>
<td>H ST</td>
<td>standard</td>
<td>sandwich</td>
<td>sandwich</td>
<td></td>
</tr>
<tr>
<td>2 Attendance mode</td>
<td>H PF</td>
<td>standard</td>
<td>full time</td>
<td>full time</td>
<td></td>
</tr>
<tr>
<td>2 Attendance mode</td>
<td>H PT</td>
<td>standard</td>
<td>part time</td>
<td>part time</td>
<td></td>
</tr>
<tr>
<td>2 Attendance mode</td>
<td>H WF</td>
<td>standard</td>
<td>writing</td>
<td>other</td>
<td></td>
</tr>
<tr>
<td>2 Attendance mode</td>
<td>H WP</td>
<td>standard</td>
<td>writing</td>
<td>other</td>
<td></td>
</tr>
<tr>
<td>2 Attendance mode</td>
<td>H YP</td>
<td>standard</td>
<td>year out</td>
<td>other</td>
<td></td>
</tr>
<tr>
<td>3 Mode of Attendance</td>
<td>Dormant, previously full-time</td>
<td>standard</td>
<td>dormant</td>
<td>dormant</td>
<td></td>
</tr>
<tr>
<td>3 Mode of Attendance</td>
<td>Dormant, previously part-time</td>
<td>standard</td>
<td>dormant</td>
<td>dormant</td>
<td></td>
</tr>
<tr>
<td>3 Mode of Attendance</td>
<td>Full time</td>
<td>standard</td>
<td>full time</td>
<td>full time</td>
<td></td>
</tr>
<tr>
<td>3 Mode of Attendance</td>
<td>Full Time Research Student</td>
<td>standard</td>
<td>full time</td>
<td>full time</td>
<td></td>
</tr>
<tr>
<td>3 Mode of Attendance</td>
<td>Full Time/Sandwich (applicants only)</td>
<td>standard</td>
<td>sandwich</td>
<td>sandwich</td>
<td></td>
</tr>
<tr>
<td>3 Mode of Attendance</td>
<td>Open/distance learning</td>
<td>standard</td>
<td>distance learning</td>
<td>distance learning</td>
<td></td>
</tr>
<tr>
<td>3 Mode of Attendance</td>
<td>Optional year out - study related eg Erasmus</td>
<td>standard</td>
<td>year out</td>
<td>other</td>
<td></td>
</tr>
<tr>
<td>3 Mode of Attendance</td>
<td>Part time</td>
<td>standard</td>
<td>part time</td>
<td>part time</td>
<td></td>
</tr>
<tr>
<td>3 Mode of Attendance</td>
<td>Part Time Research Student</td>
<td>standard</td>
<td>part time</td>
<td>part time</td>
<td></td>
</tr>
<tr>
<td>3 Mode of Attendance</td>
<td>Sandwich</td>
<td>standard</td>
<td>sandwich</td>
<td>sandwich</td>
<td></td>
</tr>
<tr>
<td>3 Mode of Attendance</td>
<td>Short Full time (&gt; 4 weeks, &lt; 24 weeks)</td>
<td>standard</td>
<td>other</td>
<td>other</td>
<td></td>
</tr>
<tr>
<td>3 Mode of Attendance</td>
<td>Staff Research Student</td>
<td>standard</td>
<td>staff</td>
<td>other</td>
<td></td>
</tr>
<tr>
<td>3 Mode of Attendance</td>
<td>Student Midwife</td>
<td>standard</td>
<td>other</td>
<td>other</td>
<td></td>
</tr>
<tr>
<td>3 Mode of Attendance</td>
<td>Student Nurse</td>
<td>standard</td>
<td>other</td>
<td>other</td>
<td></td>
</tr>
<tr>
<td>3 Mode of Attendance</td>
<td>Student ODP</td>
<td>standard</td>
<td>other</td>
<td>other</td>
<td></td>
</tr>
<tr>
<td>3 Mode of Attendance</td>
<td>Student Physiotherapist</td>
<td>standard</td>
<td>other</td>
<td>other</td>
<td></td>
</tr>
</tbody>
</table>
WHAT'S NEXT?
business needs do you have that you think could be answered with your data?

benefits will you get from connection with other people’s data?

format is your data currently in?

format do you need your data to be in?

legal/moral/ethical implications are there of sharing?
Why: do you want to share data?

do you want to share this particular data with these particular people?
When is the right time to start sharing?
Will your data be ready for sharing?
Does the data need to be available by?
can you get your data out of your systems?

can you prepare your data for sharing?
is your data?
does your data need to be stored?
will your data be available once shared?
Who:
can currently access your data?
owns the rights to your data?
can give you permission for sharing?
do you want to share your data with?
has the expertise to work with your data?
has done similar work?
Who can help?
Bethan Ruddock
content development and project management, Digital Resources

bethan.ruddock@jisc.ac.uk
@bethanar
jisc.ac.uk