UK Data Service Discover: Visible Connections and a Structuralist Approach to Discovery/Making Data Visible - Building an Enterprise Search Solution from the Ground Up

Lucy Bell, Matthew Brumpton

UK Data Archive

IASSIST 2013, Cologne
29 May 2013
Discover: visible data

- The vision, background and context (Lucy)
- The metadata challenges and solutions (Lucy)
- The technology (Matthew)
Discover: the vision

• 2010 – 2013: review, make recommendations for and then develop a new approach to resource discovery for the UK Data Service

• the result:
  • Discover – a faceted search and browse application, which capitalises on the rich, DDI metadata, and controlled vocabularies, used to index data collections, case studies, support guides and publications
Discover: the background

• The review asked how researchers find data:
  
  • trends in information-seeking behaviour showed that users prefer simple, Google-like interfaces…
  
  • …but which still return acutely-focused and highly-relevant results
  
  • the look and feel of the interfaces should be simple but the results must achieve academic rigour
  
  • so metadata are required which are both extremely powerful AND increasingly invisible
Discover: the wider context

- Functional Requirements for Bibliographic Records (FRBR) takes a holistic, structuralist view of metadata
- All items are seen in the context of the whole, with their connections and relationships used to navigate the world they, as a group, describe
- We have applied these principles to Discover, through which users may ‘discover’:

The resource discovery can take place through a simple search, a faceted browse, or by navigating the resources’ relationships. More navigations are planned.
Discover: the local context

- Previously we had a federated search environment

- with 21 interfaces/applications
Discover: the solution

- a single search interface which allows the user to move seamlessly from one type of resource to another
  - via faceted browsing
  - and directly from within each resource type

- this required **cross-referencing** data collections with:
  - publications
  - research outputs
  - support guides
  - case studies
  - (and more cross-references are planned)

- AND the use of many **controlled vocabularies**
Discover: the metadata challenges

- facets needing controlled vocabularies
- some were already in a fit state
  - Country <nation>
  - Depositor <depositor> (forthcoming)
- others needed mapping to high levels:
  - Subject categories <topcClas>
  - Access <specPerm>
- many were populated with free text:
  - Analysis unit <anlyUnit>
  - Spatial unit <geogUnit>
  - Kind of data <dataKind>
Discover: the metadata challenges (2)

• the principles:
  • manipulate and compare metadata per element, using Google Refine and Excel
  • map to existing CV where possible (e.g. spatial units list created under a separate project, U.Geo; or DDI <dataKind> CVG’s list)
  • in some cases, create a CV
  • re-import ‘clean’ metadata to the indexes
  • CVs to be used in the future

• Over summer 2012 it took 2 staff members, working c.0.4 FTE 4 months to clean 3 elements
Discover: the challenges (3)

• the mappings:

<table>
<thead>
<tr>
<th>Facet</th>
<th>DDI element</th>
<th>Existing unique values</th>
<th>Mapped to a CV of</th>
</tr>
</thead>
<tbody>
<tr>
<td>Access</td>
<td>&lt;specPerm&gt;</td>
<td>30</td>
<td>9</td>
</tr>
<tr>
<td>Subject</td>
<td>&lt;topcClas&gt;</td>
<td>116</td>
<td>21 (TTs)</td>
</tr>
<tr>
<td>Spatial unit</td>
<td>&lt;geogUnit&gt;</td>
<td>653</td>
<td>194</td>
</tr>
<tr>
<td>Analysis unit</td>
<td>&lt;anlyUnit&gt;</td>
<td>183</td>
<td>11</td>
</tr>
<tr>
<td>Kind of data</td>
<td>&lt;dataKind&gt;</td>
<td>294</td>
<td>7</td>
</tr>
</tbody>
</table>

• plus a new CV:
  • Data type (longitudinal, qualitative/mixed, microdata etc.)

• ideas of more to come….
  • Method of data collection <collMode>
  • Method of access (Explore online or download)
Discover: the technology

Data
- MS SQL
- XML (Derived from DDI 2.1)
- ESRC OAI Feeds

Search Engine
- Solr 4.1
- Tomcat 7
- Java 7
- Windows 2008 R2

Web Application
- HTML, JavaScript, CSS
- .NET 4.0
- MS SQL
Discover: the data

- Data types
  - Date, string, integer
  - Facet, multi-value
- Shared meta data
  - Mapping
- Cleansing
  - Dates
  - HTML / XML
Discover: Solr

• Open source solution written in Java

• Configured by XML

• Features include:
  • Full-text search
  • Faceted search
  • Highlighting
  • Spell check
  • Rich document handling

• Daily index

• Windows servers

• Challenges
  • Open source
  • Up skilling

Cloud configuration (4 production servers)
Discover: the application

• Supports a wide range of browsers and devices
  • Twitter Bootstrap - Grid system and responsive design
  • Modernizer - Feature detection

• Warning!
  • Responsive design only hides content
  • Dedicated mobile pages required to reduce load

• Page load times under 1 second
  • Minimise the data being returned
  • Use of asynchronous requests
Discover: the result

Simple searches, plus suggestions and help

Or browsing using facets
Discover: the result (2)

Immediate access to the download system/Explore online options

Navigation via the resources themselves to related items

And to user documentation

And to all the metadata via accordions or by looking at a single ‘target’ page

And variables…

Discover: Variable and Question Bank
Discover: the result (2)

Case study - research

Does being left-handed or right-handed affect academic ability?

Author: Christopher Cheyne, University of Liverpool; Marta Garcia-Friman, University of Liverpool; Neil Roberts, University of Edinburgh; Tim J. Crow, University of Oxford; and Stuart J. Leask, University of Nottingham.

Date: 5 November 2011

Type of case study: Research

About the research

You are a left-handed girl. Does that predict better grades than a right-handed boy?

Stories and superstitions about lefties’ have been around for generations. In recent years, there has been a special curiosity about whether there are links between handedness and academic abilities. In this study the authors investigated the effects of gender, writing hand, right-handedness, and UK region on mathematics and reading test scores from pupils across Britain.

To do this, the researchers analysed responses from 11-year-olds included in the National Child Development Study to find associations between cognitive ability (reading and maths scores) and a range of factors including handedness and gender.

Their findings indicate that children who write with one hand, while being better skilled with the other hand, performed worse in both the reading and maths tests than children who write with the hand which has the most hand skill, irrespective of whether the child was left- or right-handed.

This suggests that handedness tasks should be considered in conjunction with a measure of hand skill in future studies of investigating cognitive ability and handedness.

About the data

The research relied on data from the National Child Development Study (NCDS), a continuing longitudinal study that follows the lives of all those living in Great Britain who were born in 1958.
Discover: the result (3)

Variable and question bank

Search and browse variables and questions from survey datasets.

Search our variables

Reset filters | Clear search | Search only question text/responses | Auto-complete | Help

The UK Data Service is funded by the ESRC with contributions from the University of Essex and the University of Manchester.

© Copyright 2012-2013 University of Essex and University of Manchester. All rights reserved.
Questions

• [http://discover.ukdataservice.ac.uk](http://discover.ukdataservice.ac.uk)