The BRUCE Project
(Brunel Research Under a CERIF Environment)

Lorna Mitchell, Rosa Scoble & Richard Jones
Brunel University & Cottage Labs
Contents

– Introduction
  • Why BRUCE?
– BRUCE Project
  • The original aims & the final outcomes
– The CERIF Data Model
– Demonstration of SolrEyes
– Next Steps / Questions
Some thank you’s:

- Andy Youell, HESA
- Daniel Hook at Symplectic Ltd
- Project Team:
  - Richard Jones
  - Patricia Donaghy
  - Bethan Adams, Lawrence Jones, Giuseppe Sollazzo
Why BRUCE?

- Existing publications database (BRAD) and repository (BURA)

- Existing links between the Planning Office (managing the research information) and Library

- Keen to improve the management of research information but resources not available to implement a full CRIS
• JISC Funding Call in October 2010

• JISC Infrastructure for Education and Research Programme

• “projects that use CERIF to improve the interoperability of research information at one or more higher education institutions”
The BRUCE Project aimed to:

- develop a prototype tool, based on CERIF, to facilitate the analysis and reporting of research information from existing data sources

- 6 month project, funded by JISC

Project partners:

- St George’s University of London
- Cottage Labs (i.e. Richard Jones!)
- Symplectic Ltd
Planned Project Outputs

- BRUCE and CERIF
  - Mapping the HESA data to CERIF
  - Gap analysis

- Prototype tool
  - Born to Run on Apache Solr and Project Blacklight
  - Open Source Software

- Sample Reports
BRUCE Reports

– Bring together research information from different sources:
  • Contract information
  • Academic areas
  • PhD students supervised
  • Publications
  • Funding
  • etc.
<table>
<thead>
<tr>
<th>Subject</th>
<th>REF Eligible</th>
<th>Position</th>
<th>Full Name</th>
<th>Gender</th>
<th>FT/PT</th>
<th>Contract End Date</th>
<th>Funding Code</th>
<th>FTE</th>
<th>PhD Sup.</th>
<th>Journal Articles</th>
<th>Books</th>
<th>Chapters</th>
<th>Conf.</th>
<th>Exh./Perf./PaR</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Yes</td>
<td>Professor</td>
<td>AAA</td>
<td>F</td>
<td>FT</td>
<td></td>
<td>1.00</td>
<td>3</td>
<td></td>
<td>6</td>
<td>1</td>
<td>3</td>
<td>8</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>BBB</td>
<td>M</td>
<td>FT</td>
<td></td>
<td>1.00</td>
<td>2</td>
<td></td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CCC</td>
<td>M</td>
<td>PT</td>
<td></td>
<td>0.80</td>
<td>1</td>
<td></td>
<td>8</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>DDD</td>
<td>M</td>
<td>FT</td>
<td></td>
<td>1.00</td>
<td>1</td>
<td></td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Reader</td>
<td>XXX</td>
<td>M</td>
<td>FT</td>
<td></td>
<td>1.00</td>
<td>2</td>
<td></td>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>FFF</td>
<td>M</td>
<td>FT</td>
<td></td>
<td>1.00</td>
<td>1</td>
<td></td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Senior Lect.</td>
<td>TTT</td>
<td>F</td>
<td>FT</td>
<td></td>
<td>1.00</td>
<td>1</td>
<td></td>
<td>12</td>
<td></td>
<td></td>
<td></td>
<td>5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>YYY</td>
<td>M</td>
<td>FT</td>
<td></td>
<td>1.00</td>
<td>1</td>
<td></td>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>UUU</td>
<td>F</td>
<td>FT</td>
<td></td>
<td>1.00</td>
<td>1</td>
<td></td>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ZZZ</td>
<td>M</td>
<td>FT</td>
<td></td>
<td>1.00</td>
<td>1</td>
<td></td>
<td>5</td>
<td></td>
<td></td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Lecturer</td>
<td>EEE</td>
<td>M</td>
<td>FT</td>
<td></td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>GGG</td>
<td>M</td>
<td>PT</td>
<td></td>
<td>0.60</td>
<td>1</td>
<td></td>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td>5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>HHH</td>
<td>F</td>
<td>FT</td>
<td></td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Research</td>
<td>III</td>
<td>M</td>
<td>FT</td>
<td>31/06/2012</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Yes Total</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>13.40</td>
<td>15</td>
<td></td>
<td>88</td>
<td>1</td>
<td>11</td>
<td>25</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>Lecturer</td>
<td>JJJ</td>
<td>M</td>
<td>FT</td>
<td>16/07/2013</td>
<td>EPSRC01</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>KKK</td>
<td>M</td>
<td>FT</td>
<td>13/07/2013</td>
<td>ESRC06</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>LLL</td>
<td>M</td>
<td>PT</td>
<td>31/08/2012</td>
<td></td>
<td>0.20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Research</td>
<td>MMM</td>
<td>M</td>
<td>FT</td>
<td>31/05/2013</td>
<td>INDUS04</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>NNN</td>
<td>M</td>
<td>PT</td>
<td>31/07/2013</td>
<td></td>
<td>0.50</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>No Total</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3.70</td>
<td>0</td>
<td></td>
<td>13</td>
<td>0</td>
<td>1</td>
<td>10</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Business &amp; Management Total</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>17.10</td>
<td>15</td>
<td></td>
<td>101</td>
<td>1</td>
<td>12</td>
<td>35</td>
<td>6</td>
</tr>
</tbody>
</table>
So, what have we actually done...
The Project Outputs

– BRUCE and CERIF
  • Created a new data model for CERIF
  • Created a test dataset for CERIF
  • Met with euroCRIS to discuss gaps identified in the CERIF schema

– Prototype tool
  • Based on Apache Solr
  • Open Source Software - SolrEyes
Data model for CERIF

- Used the sample reports to identify the tables to be mapped to CERIF
- Data modelled using a spreadsheet model of the CERIF schema and then imported into a MySQL database
Using CERIF – issues:

• CERIF is predominantly relational
  – Very flexible
  – Semantics took a while to understand
  – Some unnecessary duplication required

• Publications data proved particularly tricky
  – Different field sizes
  – Date formats needed tweaking
  – Dummy defaults required for null values
• Solr (i.e. the technical bit!)
  – Using the CERIF data model the data from existing systems is mapped into a MySQL CERIF database
  – A set of Solr configuration files and data importers were then created to relate the data to a set of flat key/value pairs
  – The data is then displayed in SolrEyes
Switching off the Blacklight:

- Original plan was to use Blacklight as the user interface
  - Installation process was unstable
    - Ubuntu was difficult, although not impossible (How-To guide patched some holes in existing online guides)
    - Results on Windows were variable
    - Installation on Mac OS X proved too error prone to complete at all
  - Ranged faceting not supported in default install
• SolrEyes
Next Steps:

• Funding for current project ends on 30/09/11
  – Processes will be documented on the project blog
  – Software and supporting documentation will be made available
  – Final project reports will be published via JISC

• Investigate sources of funding to take the project forwards
Future Work:

• Further analysis of data structures & mappings
• Automated data extraction
• Development of a more sophisticated reporting component
Any Questions?

• Contacts:
  – BRUCE Project Blog
    • http://bruceatbrunel.wordpress.com/
  – Email
    • Lorna.Mitchell@brunel.ac.uk
    • Rosa.Scoble@brunel.ac.uk