Consortium of University Research Libraries
Research Libraries of the British Isles in co-operation

CURL / RSLP Collection Mapping Project
OCLC Lacey iCAS Software

Marie-Pierre Détraz
CURL Executive Secretary
Collection Mapping – What For?

• Collection management at institution’s level
• Resource Sharing / Collaborative collecting agreements: Who has what?
• Access (CURL access policy & Research Libraries Plus)
• Inter-lending (BLDSC can’t do it all)
Collection Mapping – How to do it?

- Manually – not all catalogues have been automated yet
- Problems with the manual approach: labour-intensive and high level of subjectivity
- Automated:
  - Reliability of data?
  - Level of objectivity and sophistication?
The OCLC Lacey iCAS Software

- Identifies machine-readable records on the basis of subject classification
- Supports LC, Dewey & NLM
- Maps classification to the WLN LC Conspectus scheme
- Performs analyses of individual collections & cross-institutional overlap and uniqueness analyses (tables & graph views)
The UK Context

- Many research libraries do not use or haven’t always used standard classification schemes
- Is there a way of identifying the subjects of items described in records without LC, Dewey or NLM classification numbers, so they can be included in the analysis of the collection?
The CURL / RSLP Project

• Costs:
  – £52,695 to OCLC
  – Staff time in the 6 partner institutions
  – Staff time at CURL

• Funding:
  – £28,000 from RSLP / the rest from CURL
  – Staff time paid for by the partner institutions
  – Staff time paid by CURL
Partner Institutions

- Edinburgh University Library (CURL)
- Hull University Library
- Imperial College Central Library (CURL)
- Liverpool University Library (CURL)
- Natural History Museum Library
- School of Oriental and African Studies Library
Deliverables

- An individual collection analysis for each partner library on CD-ROM at division/category/subject/title level over the years: pre-1500, per century & 20 century onwards per decades
- A combined analysis of cross-institutional overlap and uniqueness at division/category/subject level
- An external evaluator’s report
- Final report to be submitted by September 02
OCLC Methodology

- Libraries completed the Planning Guide & Questionnaire: classification methods and standards used, local idiosyncrasies
- Data sent to OCLC via ftp
- ‘No Call Number’ records matched:
  - against the other partner libraries’ files
  - against WorldCat
Results – All Partner Libraries

- Number of book format records sent: 2,767,669
- Number of deduplicated records: 2,707,696
- Number of records analysed: 2,268,225, ie 84% of the total of deduplicated records

- Number of call numbers produced at the results of the 2 matching processes: 570,705
- Method has increased by ca. 20% the number of records that could be analysed
## Results – Per Institution

<table>
<thead>
<tr>
<th>Institution</th>
<th>No of records analysed</th>
<th>% deduplicated records</th>
</tr>
</thead>
<tbody>
<tr>
<td>Edinburgh</td>
<td>604,531</td>
<td>78</td>
</tr>
<tr>
<td>Hull</td>
<td>451,590</td>
<td>99.9</td>
</tr>
<tr>
<td>Imperial College</td>
<td>202,758</td>
<td>74</td>
</tr>
<tr>
<td>Liverpool</td>
<td>536,816</td>
<td>85</td>
</tr>
<tr>
<td>NHM</td>
<td>45,820</td>
<td>47</td>
</tr>
<tr>
<td>SOAS</td>
<td>426,710</td>
<td>92</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>2,268,225</strong></td>
<td><strong>84</strong></td>
</tr>
</tbody>
</table>
Results – Per Institution

<table>
<thead>
<tr>
<th>Institution</th>
<th>No of records analysed &amp; % of deduplicated records</th>
<th>No of call numbers after matching &amp; %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Edinburgh</td>
<td>604,531 78% (-22)</td>
<td>269,468 45%</td>
</tr>
<tr>
<td>Hull</td>
<td>451,590 99.9% (-0)</td>
<td>11 0%</td>
</tr>
<tr>
<td>Imperial Col.</td>
<td>202,758 74% (-26)</td>
<td>101,105 50%</td>
</tr>
<tr>
<td>Liverpool</td>
<td>536,816 85% (15)</td>
<td>131,757 25%</td>
</tr>
<tr>
<td>NHM</td>
<td>45,820 47% (-53)</td>
<td>44,345 97%</td>
</tr>
<tr>
<td>SOAS</td>
<td>426,710 92% (-8)</td>
<td>24,019 6%</td>
</tr>
<tr>
<td>Total</td>
<td>2,268,225 84% (-16)</td>
<td>570,705 25%</td>
</tr>
</tbody>
</table>
Reliability of Data (1)

• 3 libraries – Edinburgh, Hull & Liverpool – satisfied that the data provide a fair analysis of their collections

• SOAS identified a number of inaccuracies (e.g. Chinese history classified as Italian history), caused by some idiosyncrasies of their class-marks OCLC was unaware of; data have been reloaded
Reliability of Data (2)

- The NHM were disappointed that their monographic serials weren’t included in the analysis – caused by a failure of communication, rather than a software failure.
- Both the NHM and Imperial College voiced deep reservations about the record matching process and the Conspectus scheme.
- Also some concerns about the overlap / uniqueness analysis.
Record Matching Process & Conspectus

• Inconsistencies in classification:
  – a specialised research library is likely to choose the most precise class mark, but a record from WorldCat may well contain the more general one;
  – LC and Dewey do not treat all subjects in the same way, but the software selects whichever call number is available in the matching record;

• Conspectus classification: ‘blunt’, weak in certain subjects and outdated
### Uniqueness Analysis (all) - 1

<table>
<thead>
<tr>
<th>Library</th>
<th>Unique (incl. no call numbers) From OCLC data</th>
<th>% of all records (my own calculations)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Edinburgh</td>
<td>586,662</td>
<td>75%</td>
</tr>
<tr>
<td>Hull</td>
<td>334,893</td>
<td>71%</td>
</tr>
<tr>
<td>Imperial Col.</td>
<td>199,412</td>
<td>72%</td>
</tr>
<tr>
<td>Liverpool</td>
<td>491,371</td>
<td>74%</td>
</tr>
<tr>
<td>NHM</td>
<td>88,252</td>
<td>88%</td>
</tr>
<tr>
<td>SOAS</td>
<td>426,109</td>
<td>90%</td>
</tr>
<tr>
<td>Total</td>
<td>2,126,699</td>
<td>77%</td>
</tr>
</tbody>
</table>
Uniqueness Analysis (all) - 2

Flawed?

- Not in WorldCat = unique?
- Need to analyse the no call numbers
- Need to assess uniqueness on the basis of the analysed records (not in the scope of this study?)
- Need clarification from OCLC on the methodology used for uniqueness and overlap analysis
Overall Evaluation

• **Plus Points:**
  – Number of records analysed increased by ca. 20%
  – A fair analysis of the individual collections

• **Minus Points:**
  – Inconsistencies in classification
  – Records with no call numbers might represent what is the most unique in the collection

• **Question Marks:**
  – Reliability of overlap / uniqueness analysis?
  – Will it be possible to analyse overlap and uniqueness at title level?