

***curl*** 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

***curl***

Consortium of University Research Libraries

Research Libraries of the British Isles in co-operation

## **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



Consortium of University Research Libraries

Research Libraries of the British Isles in co-operation

## 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



Consortium of University Research Libraries

Research Libraries of the British Isles in co-operation

## Results – Per Institution

	No of records analysed	% deduplicated records
Edinburgh	604,531	78
Hull	451,590	99.9
Imperial College	202,758	74
Liverpool	536,816	85
NHM	45,820	47
SOAS	426,710	92
Total	2,268,225	84



Consortium of University Research Libraries

Research Libraries of the British Isles in co-operation

## Results – Per Institution

No of records analysed &  
% of deduplicated records

No of call numbers  
after matching & %

<b>Edinburgh</b>	<b>604,531</b>	<b>78% (-22)</b>	<b>269,468</b>	<b>45%</b>
Hull	451,590	99.9% (-0)	11	0%
<b>Imperial Col.</b>	<b>202,758</b>	<b>74% (-26)</b>	<b>101,105</b>	<b>50%</b>
Liverpool	536,816	85% (15)	131,757	25%
<b>NHM</b>	<b>45,820</b>	<b>47% (-53)</b>	<b>44,345</b>	<b>97%</b>
SOAS	426,710	92% (-8)	24,019	6%
Total	2,268,225	84% (-16)	570,705	25%

## 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



Consortium of University Research Libraries

Research Libraries of the British Isles in co-operation

## Uniqueness Analysis (all) - 1

	Unique (incl. no call numbers) From OCLC data	% of all records (my own calculations)
Edinburgh	586,662	75%
Hull	334,893	71%
Imperial Col.	199,412	72%
Liverpool	491,371	74%
NHM	88,252	88%
SOAS	426,109	90%
Total	2, 126, 699	77%

## 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?