

**CURL / RSLP COLLECTION MAPPING PROJECT**

**based on**

**OCLC / LACEY iCAS SOFTWARE**

**FINAL REPORT**

**September 2002**

## **Executive Summary**

The OCLC / Lacey iCAS software provides an automated approach to the analysis of library collections by identifying bibliographic records in library online catalogues on the basis of subject classification. The software picks up the classification number that is included in the bibliographic record, and classification is then automatically mapped to the WLN / LC Conspectus subject headings.

However, the iCAS software supports only three standard schemes (Library of Congress (LC), Dewey and the National Library of Medicine scheme), while UK libraries use a variety of classification systems – some of them in-house. The purpose of this project was to explore ways of adapting the software in such a way that it would be able to analyse variously classified collections, both at institutional and cross-institutional level, in a UK context.

Six institutions took part in the project: the libraries of the universities of Edinburgh, Liverpool and Hull, which hold collections across a broad range of subjects; and the libraries of Imperial College, the Natural History Museum and the School of Oriental and African Studies, which hold more specialised collections. The collections analysed in the framework of this project were, for the most part, print collections, as described in the partner libraries' online catalogues.

The project has demonstrated that (a) the iCAS software can be adapted to a UK context – the methodology developed by OCLC to identify records without LC or Dewey classification numbers increased the number of records that could be analysed across the six partner libraries from 63% to 84%, i.e. by 21%; (b) the iCAS software can provide a fair, albeit incomplete and relatively blunt, analysis of collections described in online catalogues in terms of Conspectus categories.

The project has also highlighted the need:

- (a) To refine the iCAS software with a view to ensuring a higher percentage of records analysed - the proportion of records that could not be analysed ranges from 0.1% to 53% across the six partner libraries – as well as a more sophisticated and up-to-date set of Conspectus subject headings;

If this is the route chosen, a number of actions would need to be taken, including:

- the partner libraries to examine their 'NO CALL NUMBER PRESENT' files to find out why these records could not be analysed by the software in spite of the method devised by OCLC;
- the overlap and uniqueness analyses to be re-run without the 'NO CALL NUMBER PRESENT' files, as it cannot be assumed that these records describe unique items;
- the partner libraries – and/or possibly CURL libraries in general – to work with OCLC on ways of improving and updating the Conspectus system of subject headings;
- the partner libraries to test the web-based application of the software being developed by OCLC, with a view, in particular, to testing the uniqueness and overlap analyses at title level, which is not possible with the current software;

It would also be extremely worthwhile to combine the work aimed at refining the iCAS software with a re-run of the data of the partner libraries whose catalogues

have been much improved by the addition of records produced in the framework of the RSLP and other externally funded projects.

(b) To explore the financial implications of an extended use of the software across the Consortium and/or across the UK for a possible nationwide collaborative collection management exercise;

(c) To discuss the outcomes of this project with relevant UK research centres and services providers, such as UKOLN, The Centre for Digital Library Research and MIMAS, with a view to comparing the OCLC approach with other automated approaches in order to inform further work at national level;

## **Contents**

### **1 Purpose of the Project**

- 1.1 *Collection Mapping*
- 1.2 *The OCLC / Lacey iCAS Software*
- 1.3 *The Project*
- 1.4 *The Partnership*
- 1.5 *The Final Report*

### **2 Methodology and Results**

- 2.1 *Methodology*
- 2.2 *Results*
  - 2.2.1 Cross-institutional results
  - 2.2.2 Institutional results
  - 2.2.3 Matching Results

### **3 Evaluation of Results**

- 3.1 *Evaluation of the experiment*
  - 3.1.1 Positive outcomes
  - 3.1.2 Hiccups
  - 3.1.3 Problems
  - 3.1.4 Reservations
- 3.2 *Evaluation of the software*
  - 3.2.1 Plus points
  - 3.2.2 Minus points

### **4 Conclusion and Recommendations**

## **5 Appendices**

5.1 *RSLP Circular 2000/3*

5.2 *CURL Response to RSLP 2000/3*

5.3 *CURL Project Proposal*

5.4 *Letter of Agreement between CURL and OCLC*

5.5 *OCLC Guide to Using the Interactive Collection Analysis System*

5.6 *Project Summary by Sally Loken, OCLC ACAS Consultant*

5.7 *Additional Data from OCLC*

5.8 *Responses from Partner Libraries (with, in some cases, replies from OCLC)*

5.9 *Example of Statistical Tables and Graph Views (Uniqueness and Overlap)*

5.10 *What We Learned – Suggested Improvements for Future Analysis, from OCLC*

5.11 *External Evaluation Report by Dennis Nicholson, Director of the Centre for Digital Library Research, University of Strathclyde*

5.12 *Acronyms*

5.13 *Key contacts*

# 1 Purpose of the Project

## 1.1 Collection Mapping

CURL and RSLP share a common goal in their effort to encourage institutions both in the HE and public sectors – the three national libraries joined CURL as full members in October 02 - to work collaboratively with a view to enhancing access to research collections for all researchers across the UK. In order to achieve this goal, which is also very much in accord with the focus of the Research Support libraries Group (RSLG), there is a need to develop new models of effective resource sharing based on a national collection mapping strategy.

The systematic description of library collections at national level and across the sectors is absolutely crucial in order:

- (a) To provide effective and reliable guidance for researchers (e.g. user guides identifying strong collections in particular subjects with link to appropriate online catalogues; cross-searching for distributed collections etc.); this is particularly relevant in the context of the CURL access policy and the Research Libraries Plus initiative;
- (b) To describe the strengths and weaknesses of the various depositories with a view, in particular, to identifying collaborative collecting and preserving responsibilities and distributed inter-lending models; this is particularly important in the context of the Full Disclosure programme, whose ultimate goal is to implement the automation of the catalogues of all major research collections held in the UK;

Although libraries have traditionally focused on item-level description, a number of different approaches to collection-level description have emerged over the past few years (e.g. Conspectus (RLG), Dublin Core or more recently the RSLP collection mapping web tool designed by UKOLN). However all these approaches have one thing in common – they all leave the actual description of the collection to the creator of the record, which can result in a significant element of subjectivity in the way the strengths and weaknesses of the collections are described. There is also the fact that manual assessments are extremely labour-intensive.

This is why CURL and RSLP became interested in the potential of a software tool developed by WLN, now OCLC / Lacey, which provides an automated approach to analysis of library collections, and agreed to fund a pilot that would provide:

- (a) each partner library with a useful snapshot of its collections as described in its online catalogue;
- (b) at a consortial level, an overlap / uniqueness analysis across all the partner libraries;

## **1.2 The OCLC / Lacey iCAS Software**

The OCLC / Lacey iCAS software provides an automated approach to the analysis of library collections by identifying bibliographic records in library online catalogues on the basis of subject classification. The software picks up the classification number (also referred to as class-mark or shelf-mark) that is included in the bibliographic record, and classification is then automatically mapped to the WLN / LC Conspectus subject headings.

However there is a variety of classification schemes in use – some of them in-house – and the iCAS software supports only the three main standard schemes: Library of Congress (LC), Dewey and the National Library of Medicine (NLM) scheme. While this is perfectly appropriate for most research libraries in the US, where the software is already in use, this poses a problem in the UK where many CURL and other research libraries either do not use or haven't always used these schemes. CURL therefore asked OCLC whether they would be interested in exploring ways of adapting the software in such a way that it would be able to analyse variously classified collections, both at institutional and cross-institutional level, in a UK context.

## **1.3 The Project**

It was agreed that OCLC would carry out a pilot project, to be jointly funded by CURL and RSLP, aimed at analysing the machine-readable book format records of six UK research libraries using different and often non-standard classification schemes. The project's deliverables are:

- (a) An individual collection analysis for each partner library at the division/category/subject/title level of the WLN LC Conspectus over a number of years as defined by CURL (pre-1500, then per century and from the 20<sup>th</sup> century onwards per decade); each partner library has been provided with an OCLC iCAS CD-ROM disc with the analysis of its own collections; the reports consist of statistical tables and graphs views of the collections' age and content;
- (b) A combined analysis at the division/category/subject level identifying cross-institutional overlap and uniqueness; both CURL and RSLP have been provided with copies of an OCLC iCAS CD-ROM disc with the combined analysis; the reports consist of statistical tables and graph views of the collections' age, content, uniqueness and overlap.
- (c) An external evaluator's report (see Appendix 5.11) by Dennis Nicholson, Director of the Centre for Digital Library Research at the University of Strathclyde. Dennis Nicholson has been actively involved with a number of projects, notably HILT, SCONE and CAIRNS, which are all concerned with the measurement of collection strengths as a means towards collaborative collection management.
- (d) A final report (i.e. this document);

The costs of the project have amounted to £53,695 (£52,695 to OCLC and £1,000 to the consultant), to which must be added staff time in the six partner institutions, paid for by the partner institutions, and staff time at CURL, paid for by CURL.

#### **1.4 The Partnership**

Six libraries have taken part in the project:

- Three – the libraries of the universities of Edinburgh, Liverpool and Hull – represent fairly typical university libraries that hold collections across a broad range of subjects;
- Three – the libraries of Imperial College, the Natural History Museum (NHM) and the School of Oriental and African Studies (SOAS) – hold more specialised collections;

The partner libraries use a range of classification methods, including standard and in-house schemes, and in some cases a combination of two or more systems. As pointed out earlier, this reflects the reality of classification usage within CURL and more widely within the UK.

Three of the partner libraries – Edinburgh, Imperial College and Liverpool - are CURL members.

More information about the six institutions is available in the *Project Proposal* (see Appendix 5.3).

The collections analysed in the framework of the project were, for the most part, print collections, as described in the partner libraries' online catalogues.

#### **1.5 The Final Report**

This final report has been compiled on the basis of five documents or sets of documents and data produced by the various participants at different stages in the course of the project:

- (1) The *Project Summary* from OCLC, dated 14 December 01 (see Appendix 5.6) and additional data provided by OCLC while this report was being compiled (see Appendix 5.7);
- (2) The CD-ROM containing the combined data analysis;
- (3) The comments of the participating libraries following the demonstration of the software by the OCLC team on 10 January 02, with, in some cases, OCLC's responses to these comments (see Appendix 5.8);
- (4) OCLC's general response to these comments, *What we learned – Suggested Improvements for Future Analyses*, dated 3 March 02 (see Appendix 5.10);
- (5) The external evaluator's report (see Appendix 5.11): this report focuses on the broad issues pertaining to the pros and cons of an automated approach to collection mapping, rather than on the methodology experimented by the OCLC team to adapt the use of the iCAS software to a UK context. It therefore needs to be read separately as an adjunct to the final report.

## **2 Methodology and Results**

The following analysis is based on the OCLC *Project Summary* document, dated 14 December 01 (see Appendix 5.6), OCLC's additional data (see Appendix 5.7) and on the combined analysis CD-ROM.

### **2.1 Methodology**

Before sending their data to OCLC, the partner libraries completed the OCLC / Lacey Planning Guide and Questionnaire, the purpose of which was to inform the OCLC team about local convention and standards in classifying materials and where, within the MARC record, other relevant data are located.

All data were sent to OCLC by ftp. Hull, the NHM and SOAS sent their data directly, while the data of the CURL libraries (Edinburgh, Imperial College and Liverpool) were sent from Manchester Computing, which hosts the CURL database.

In order to identify the subjects of the books described in the records without LC or Dewey classification numbers – these records will be referred to as the 'no call number records' - OCLC devised the following procedure:

- (a) The 'no call number records' of each library were first matched with the files of the other partner libraries in order to obtain call numbers if present in the other libraries' records;
- (b) The remaining 'no call number records' were then matched against WorldCat, OCLC's union catalogue, which contains 47 million bibliographic records (see <http://www.oclc.org/worldcat>);

The records were matched on titles, authors, LCCNs, edition, publisher, date, material type, ISBN and ISSN.

### **2.2 Results**

This is a summary of the results based on statistics provided by OCLC.

#### **2.2.1 Cross-institutional results**

See table on next page.

Total number of books	2,767,669	
Total number of records after de-duplication of records in each library's file	2,707,696	
<b>Total number of records with call numbers (i.e. with LC or Dewey classification numbers) before the matching processes</b>	<b>1,710,142</b>	<b>i.e. 63% of de-duplicated records</b>
Total number of call numbers produced as the result of the 2 matching processes	570,885	i.e. 57% of no call number records
Total number of records with call numbers after the 2 matching processes	2,281,027	i.e. 84.24% of de-duplicated records
Total number of records rejected because of ill-formed call numbers	12,802	
<b>Total number of records analysed</b>	<b>2,268,225</b>	<b>i.e. 84% of de-duplicated records</b>

These figures show that the methodology devised by OCLC to gather records without call numbers (i.e. without LC or Dewey classification numbers):

- (a) Has increased by 21%, from 63% to 84%, the number of records that could be analysed.
- (b) Has resulted in 57% of the records that were initially without call numbers being successfully matched against the other partner libraries' files or WorldCat.

### 2.2.2 Institutional Results

Detailed information about the records of each partner library can be found on pages 2-3 of the OCLC *Project Summary* document (see Appendix 5.6). The following table contains a summary of the main results:

	Number of records after de-duplication	Number of records with call numbers before the matching processes	Number of records analysed after the matching processes (& minus the rejected records)	% of records with call numbers before the matching processes	% of records analysed after the matching processes (& minus the rejected records)	% of increase in the numbers of records analysed
Edinburgh	772,923	336,476	604,531	44	78	34
Hull	460,653	460,636	451,590	99.9	99.9	0
Imperial	274,398	102,032	202,758	37	74	37
Liverpool	634,714	405,801	536,816	64	85	21
NHM	97,590	1,528	45,820	2	47	45
SOAS	467,418	403,669	426,710	86	92	6
Total	2,707,696	1,710,142	2,268,225	63	84	21

The figures in the above table show that:

- Results vary a great deal from one institution to another;
- While Hull is the library that benefited the most from the exercise inasmuch as the software could call practically all its records even before the matching processes – obviously because Hull uses standard class-marks – it is the NHM which has benefited the most from the methodology devised by OCLC: as a result of this methodology, the number of NHM records that could be analysed increased by 45%, from 2% to 47%. The figures for Imperial (37%), Edinburgh (34%) and Liverpool (21%) also indicate a substantial increase in the number of records analysed as a result of the methodology devised by OCLC.

### 2.2.3. Matching results

As explained in section 2.1, the records with no call numbers were first matched against the files of the other partner libraries. Then the remaining records with no call numbers were matched against WorldCat. The table below shows that 76% of the records successfully matched came from WorldCat.

Library	Number of call numbers found by matching each other	% of call numbers found by matching each other	Number of call numbers found by matching WorldCat	% of call numbers found by matching WorldCat	Total
Edinburgh	59,385	22	210,263	78	269,648
Hull	5	45	6	55	11
Imperial College	29,984	30	71,121	70	101,105
Liverpool	36,439	28	95,318	72	131,757
NHM	8,800	20	35,545	80	44,345
SOAS	0	0	24,019	100	24,019
Total	134,613	24	436,272	76	570,885

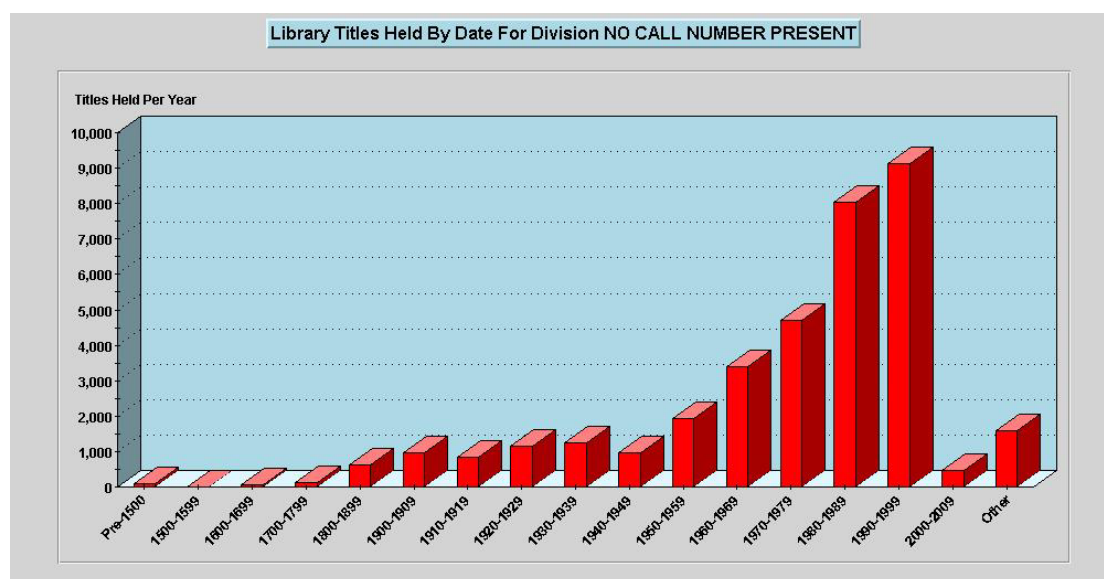
## 2.2.4 Analysis of 'no-call number records'

As shown in sections 2.2.1 and 2.2.2, 16% of all de-duplicated records could not be matched against any records whether from the combined files of the partner libraries or WorldCat. Here is a breakdown of the percentages of 'no call number records' by institution, which again shows that results vary a great deal from one institution to another:

	% of 'no call number records left after the 2 matching processes
Edinburgh	22.0
Hull	0.1
Imperial College	26.0
Liverpool	15.0
NHM	53.0
SOAS	8.0
Total	16.0

As agreed with CURL, OCLC created a division and category called NO CALL NUMBER PRESENT, available on the combined analysis CD-ROM, where the 'no call number records' can be analysed by date and holding library, yet not by subject area.

The graph from the combined analysis CD-ROM reproduced below shows a steady increase in 'no call number records' over the period 1950-1999:



The spreadsheet from the combined analysis CD-ROM reproduced below shows that 'no call numbers' over the period 1950-1999 make up 64% of all 'no call number records'.

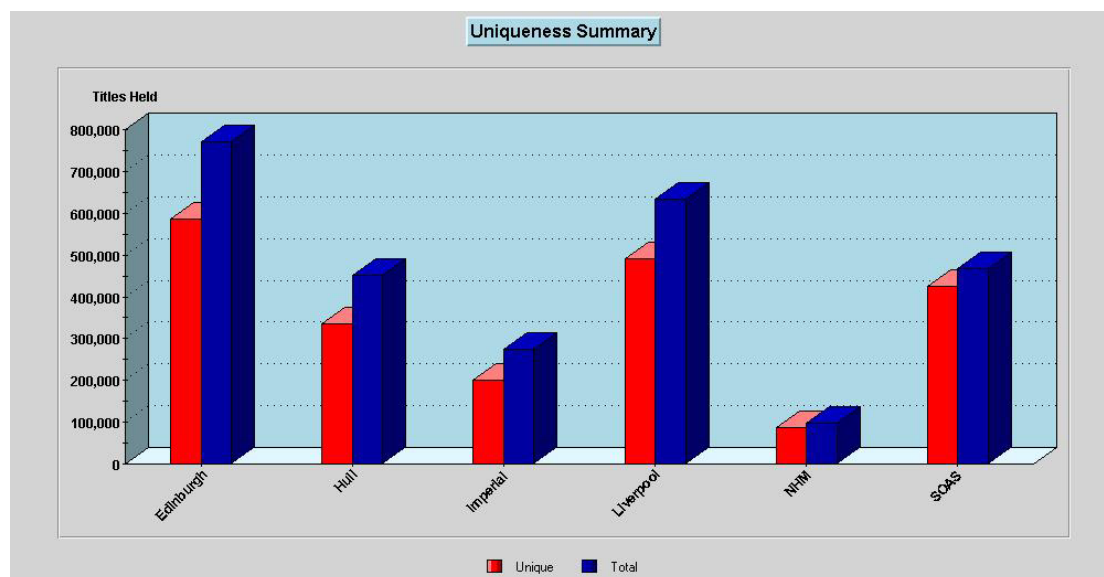
**Library Titles held by Date (Category Level, Select Div Library)**

INSTITUTION	Pre-1500	1500-1599	1600-1699	1700-1799	1800-1899	1900-1909	1910-1919	1920-1929	1930-1939	1940-1949	1950-1959	1960-1969	1970-1979	1980-1989	1990-1999	2000-2009	Other	Total
Edinburgh	3	245	2698	5337	29574	1829	1773	2466	2906	2919	6587	12581	17289	36390	39813	1042	3347	<b>166799</b>
Imperial College	0	25	79	409	3188	876	820	1151	1252	894	2303	5094	11385	17079	22231	1788	2687	<b>71261</b>
NHM	11	150	268	1039	7130	1609	1425	1692	1483	776	1366	1997	3232	8395	14966	1426	4752	<b>51717</b>
SOAS	0	32	147	226	3060	916	988	1187	1618	1333	2881	3913	4872	8143	6672	383	3359	<b>39730</b>
Hull	0	0	0	0	0	0	0	0	0	0	1	0	1	0	4	0	0	<b>6</b>
Liverpool	94	918	3183	6958	17851	4351	3437	5040	5303	3634	6148	10373	10297	10414	7500	0	1655	<b>97156</b>
<b>Grand Totals</b>	<b>108</b>	<b>1370</b>	<b>6375</b>	<b>13969</b>	<b>60803</b>	<b>9581</b>	<b>8443</b>	<b>11536</b>	<b>12562</b>	<b>9556</b>	<b>19286</b>	<b>33958</b>	<b>47076</b>	<b>80421</b>	<b>91186</b>	<b>4639</b>	<b>15800</b>	<b>426669</b>

**It is however unclear why these records could not be matched, as no detailed analysis of these records has been carried out by the individual libraries within the framework of this project.**

### 2.3 Overlap and Uniqueness Analysis

The combined analysis has been performed at division/category/subject level. At this stage in its development the software cannot perform an overlap and uniqueness analysis at title level. The following graph is from the combined analysis CD-ROM:



The statistics below give a more precise picture. The figures in the second column of this table have been extracted from the *Uniqueness Analysis (all)* on the combined analysis CD-ROM. The percentages in the third column of the table have been calculated manually:

INSTITUTION	Unique	% all records
Edinburgh University	586662	<b>75</b>
Imperial College	199412	<b>72</b>
Natural History Museum	88252	<b>88</b>
School of Oriental and African Studies	426109	<b>90</b>
University of Hull	334893	<b>71</b>
University of Liverpool	491371	<b>74</b>
Total	2126699	<b>77</b>

Note: for a detailed analysis of the uniqueness figures per institution and subjects, see Appendix 5.9.

It is extremely important to bear in mind, while looking at the above figures, that the overlap and uniqueness analyses contain all the records of the partner libraries, including the records that have remained without call numbers after the completion of the two matching processes. This might account for the surprisingly high percentages in the third column. With hindsight it might have been wiser to exclude the 'no call number records' from the overlap / uniqueness analyses. As pointed out in section 2.2.3, it has not been established why these records couldn't be matched against the partner libraries' files or WorldCat. It cannot be assumed that these records couldn't be matched because they describe unique items.

### 3 Evaluation of Results

The following assessment is based on:

- (a) The data analysis in Section 2
- (b) The comments sent by the partner libraries about their own institutional data following the seminar organised on 10 January 02 (see Appendix 5.8)
- (c) OCLC's response to these comments (see Appendices 5.8 and 5.10)
- (d) The external assessor's report (see Appendices 5.11)

The assessment is divided in two parts:

- Evaluation of the experiment carried out by the project
- Evaluation of the software as a whole

#### 3.1 Evaluation of the experiment

##### 3.1.1 Positive outcomes

- The methodology designed by OCLC to identify records without LC or Dewey classification numbers increased the number of records that could be analysed by 21% from 63% to 84%.
- Three libraries scored results above the 84% average: Hull with 99.9% and perhaps, even more impressively, as they had a much higher percentage of 'no call number records' to begin with, SOAS with 92% and Liverpool with 85%;
- Two libraries had ca.  $\frac{3}{4}$  of their records analysed: Edinburgh (78%) and Imperial College (74%);
- Edinburgh, Hull and Liverpool were highly satisfied with the results for their institutions. They clearly stated in their comments that the data provided a fair analysis of the collections that are described on their online catalogues in terms of Conspectus categories. The few anomalies spotted by these libraries were not seen as statistically significant. (Hull, which achieved the highest percentage of records analysed, with 99.9%, have acknowledged that some inaccuracies had occurred as a result of their deviation from the LC classification scheme in the case of country divisions, but they will be able to compile a list of their deviations in order to rectify the analysis.)

##### 3.1.2 Hiccups

- Following the demonstration of the software by the OCLC team on 10 January 02, SOAS identified a worrying number of inaccuracies in their data (e.g. Chinese history classed as Italian history). Further analysis, however, revealed that these errors stemmed from SOAS's failure to inform the OCLC team of all the idiosyncrasies of their Dewey-based class marks. OCLC agreed to resubmit the SOAS files and to rerun the overlap / uniqueness analysis.
- The NHM was disappointed by the fact that their monographic serials had not been included in the analysis. While this was merely the result of a

communication misunderstanding, it is a useful reminder of how crucial it is for the library and the service provider to have a close dialogue at the stage when specifications are being drawn and data prepared.

### 3.1.3 Problems

- When libraries have records without call numbers that cannot be matched against the partner libraries' files or WorldCat, some of their unique titles will not be reported. This has been identified as a serious problem by some partner libraries as well as OCLC.
- As shown in section 2.2.4. the records remaining without call numbers after the completion of the two matching processes constitute a significant proportion of the total number of records – especially in the case of the NHM (53%), but also in the case of Imperial College (26%) and Edinburgh (22%) – and there is therefore a need to investigate why these records could not be matched.

### 3.1.4 Reservations

A number of reservations have been voiced about the record matching process and the Conspectus scheme on which the iCAS software is based:

- There are some inconsistencies in the way the books are classified: e.g. Imperial College noticed that some titles were correctly classified under 'Numerical analysis', while others on the same subject appeared under 'Mathematical analysis, general'. This is the result of the fact that the class-marks used in the matching records originate from libraries contributing to WorldCat or less specialised partner libraries which tend to prefer more generalised class-marks. This was perceived as a real problem by Imperial College and especially the NHM, whose collections are highly specialised.
- The Conspectus-based classification tends to be rather 'blunt' and weak in certain subjects (the NHM gave the example of Biology in relation to animal and plant species). It needs updating (e.g. Edinburgh suggested that 'Artificial intelligence' be added under 'Computer Science' and 'Molecular biology' under 'Biological sciences').
- OCLC also suggested that there was some disagreement from the partner libraries about where their materials were mapped in the Conspectus scheme because LC and Dewey Classification schemes do not treat all subjects in the same way – and when the records are matched against the partner libraries' files or WorldCat, the software selects whichever call number is available in the matching record, whether it is an LC or a Dewey call number.

## **3.2 Evaluation of the software**

This section aims to give a broader evaluation of the software, yet will include some of the main points raised in the previous section.

### **3.2.1 Plus points**

- The iCAS software can be adapted to a UK context. The methodology developed by OCLC to identify records without LC or Dewey classification numbers increased the number of records that could be analysed by 21% from 63% to 84%.
- The iCAS software can provide a fair, albeit incomplete and relatively blunt, analysis of collections described in online catalogues in terms of Conspectus categories.

### **3.2.2 Minus Points**

- The method is objective only to a point. The data produced by the software are based on classification numbers rather than subject headings, and each book has only one classification number. So, for example, a book on Japanese architecture, which has been classified under 'Special types of building', will not appear under 'Architecture, Japan'.
- The method is fairly reliable, yet lacks in sophistication, in particular, because:
  - ✓ It produces some inconsistencies in the way the books are classified as a result of the fact that it is based on three classification schemes, which define subjects differently;
  - ✓ It does not help to assess the significance of a collection in relation to specific user groups (e.g. a strong collection to a researcher can be a weak collection to a learner).
- At the moment the iCAS software cannot perform a title analysis of the overlap or uniqueness data. However, OCLC are in the process of developing a web-based application that will allow this.
- The iCAS software can only provide libraries with snapshots of their collections at a given time rather than ongoing up-to-the minute information.
- The financial case for an extended use of the software in the UK remains to be made.

## 4 Conclusion and Recommendations

The project has demonstrated that the iCAS software is capable of providing reasonably accurate, albeit relatively unsophisticated automated analyses of collection strengths in a UK context.

However, the project has also highlighted the need:

- (c) To refine the iCAS software with a view to ensuring:
  - a. a higher percentage of records analysed - the proportion of records that could not be analysed ranges from 0.1% to 53% across the six partner libraries -;
  - b. a more sophisticated set of Conspectus subject headings;

If this is the route chosen, a number of actions would need to be taken, including:

- the partner libraries to examine their 'NO CALL NUMBER PRESENT' files to find out why these records could not be analysed by the software in spite of the matching processes devised by OCLC;
- the overlap and uniqueness analyses to be re-run without the 'no call number records';
- the partner libraries – and/or possibly CURL libraries in general – to work with OCLC on ways of improving and updating the Conspectus system of subject headings;
- the partner libraries to test the web-based application of the software being developed by OCLC, with a view, in particular, to testing the uniqueness and overlap analyses at title level, which is not possible with the current software;

It would also be extremely worthwhile to combine the work aimed at refining the iCAS software with a re-run of the data of the partner libraries whose catalogues have been much improved by the addition of records produced in the framework of the RSLP and other externally funded projects.

- (d) To explore the financial implications of an extended use of the software across the Consortium and/or across the UK for a possible nationwide collaborative collection management exercise;
- (e) To discuss the outcomes of this project with relevant UK research centres and services providers, such as UKOLN, The Centre for Digital Library Research and MIMAS, with a view to comparing the OCLC approach with other automated approaches in order to inform further work at national level (see also Dennis Nicholson's Recommendations in Appendix 5.11);

END OF REPORT

## 5 Appendices

### 5.1 RSLP Circular 2000/3

#### “RSLP Circular 200/3

3 August 2000

**To: Librarians/Directors of Information Services at  
HEFCE/SHEFC/HEFCW/DHFETE -funded institutions**

Dear Librarian/Director of Information Services

#### **Future RSLP activities and post-RSLP developments**

The Research Support Libraries Programme is shortly to review its portfolio of projects and studies, with a view to establishing its priorities for future funding activities. The Programme will also be considering longer-term developments and will be communicating its views to the HE funding bodies.

#### **a. Small projects and studies**

As part of this process I am writing to seek suggestions for small projects or studies that it might now be appropriate for RSLP to fund. The sum that would be associated with these activities would not normally be greater than £25K but, in a small number of cases, funding of work up to a maximum of £75K may be considered.

The suggested projects or studies should enhance or develop [the RSLP vision](#). The activities need not necessarily relate to disciplines or types of material previously targeted by the Programme.

Please note that this is an opportunity to inform the decision-making process, not a request for proposals.

.....”

## 5.2 CURL Response to RSLP 2000/3

“

Mr Ronald Milne  
Director  
Research Support Libraries Programme  
Edinburgh University Library  
George Square  
Edinburgh  
EH8 9LJ

29 August 00

Dear Ronald

### Research Support Libraries Programme: Circular 2000/3

I am responding on behalf of CURL (Consortium of University Research Libraries) to RSLP Circular 2000/3.

a) Suggestions for small projects or studies which it might be appropriate for RSLP to fund.

CURL currently has twenty-six institutions in membership, and has recently revised its strategic directions (<http://www.curl.ac.uk>). One of the most important elements concerns resource management, where CURL is seeking '*To optimise the combined information resources and expertise of the CURL libraries, in partnership with relevant organisations, for the benefit of research scholarship..*'. There are a number of areas of activity within this but one is concerned with '*..the development of a national collection mapping strategy...*'.

CURL members are involved, through RSLP and eLib-funded projects with the creation of collection level descriptions. Additionally, we have provided the data for a pilot register of collection strengths, developed by David Haynes for the NPO. CURL is also managing the infrastructure provision for the HE Archives Hub, which is and will be populated with collection level descriptions for archival content.

This has highlighted an area of concern. There are a number of differing approaches, more or less standards-based, which have been developed by different communities. Although the metadata elements within these approaches are more or less structured, the key element – the description of the collection, or indeed, which description and content format to use is left to the creator of the record. This provides serious inconsistencies in description, which results in an unbalanced view of what is available to a researcher.

CURL has recently become aware of a software tool, which offers an alternative approach. iCas has been developed by OCLC/WLN as a tool which will analyse a library collection, down to item level, and map it, using classification, to the relevant Conspectus headings. The analysis can provide a detailed subject breakdown of collections over time. More significantly, the software can also provide analyses across consortia or groups of libraries, showing areas of overlap or uniqueness.

We would suggest that RSLP fund a pilot to evaluate the potential of this software in measuring collection strengths across a small number of CURL libraries, initially. We would propose to use the CURL database as the source of data for the evaluation. The analysis would provide:

- A detailed subject comparison of holdings down to item level
- An indication of acquisition patterns over time
- An indication of unique holdings
- An indication of overlapping holdings

We are presently in discussion with Manchester Computing and OCLC/WLN with a view to providing some test data for analysis, which will hopefully help in identifying the most appropriate methodology. In the circumstances, it is not possible to provide definitive costings but we would anticipate, after discussion with OCLC, that a project involving 5 research libraries with multi-disciplinary holdings and no more than 5mll records in total would cost approximately £60,000 (exclusive of VAT). CURL would be willing to make a partnership contribution towards any study.

The benefits of such a neutral, and objective overview are significant. The results would provide a basis for collaborative collection development planning and service delivery. If successful, the outcomes could potentially be scaled up to provide an overview UK-wide.

These objectives are entirely complementary to RSLP's own vision, of facilitating the best possible arrangements for research support, in increasing the availability of information about the location of the UK's rich information resources.

.....”

### **5.3 CURL Project Proposal**

#### **Proposal to Undertake a Collection Mapping Survey and Analysis**

##### ***Introduction***

CURL (Consortium of University Research Libraries) currently has twenty-six institutions<sup>1</sup> in membership. One of the strengths of the consortium is the richness of its collections and the value these have for research and scholarship nationally and internationally. In revising its strategic directions (<http://www.curl.ac.uk>), CURL is seeking '*To optimise the combined information resources and expertise of the CURL libraries, in partnership with relevant organisations, for the benefit of research scholarship..*'. There are a number of areas of activity within this but one is concerned with '*..the development of a national collection mapping strategy...*'. We regard this as a crucial step in underpinning any national collaborative collection development and management strategies.

##### ***Collection Analysis***

CURL members are involved, through RSLP and eLib-funded projects with the creation of collection level descriptions. CURL has also provided the data for a pilot register of collection strengths, developed by David Haynes Associates for the NPO; and is managing the infrastructure provision for the HE Archives Hub, which is and will be populated with collection level descriptions for archival content.

This has highlighted an area of concern. There are a number of differing approaches, more or less standards-based, which have been developed by different communities. Although the metadata elements within these approaches are more or less structured, the key element – the description of the collection, or indeed, which description and content format to use is left to the creator of the record. This provides serious inconsistencies in description, which results in an unbalanced view of what is available to a researcher.

CURL has recently become aware of a software tool, iCas, which offers an alternative approach. iCas has been developed by OCLC/Lacey as a tool which will analyse a library collection, down to item level, and map it, using classification, to the relevant conspectus headings. The analysis can provide a detailed subject breakdown of collections over time. More significantly, the software can also provide analyses across consortia or groups of libraries, showing areas of overlap or uniqueness.

CURL, in response to RSLP Circular 2000/3, has therefore suggested that some investigation into these issues be funded. The benefits of such a neutral and objective overview are significant. The results would provide a basis for collaborative

---

<sup>1</sup> The Universities of Aberdeen, Birmingham, Cambridge, Durham, Edinburgh, Glasgow, Leeds, Liverpool, Manchester, Newcastle, Nottingham, Oxford, Sheffield, Southampton, Warwick; Imperial College, King's College London, London School of Economics, Trinity College Dublin, University College London, University of London Library; the British Library, the National Library of Scotland, the National Library of Wales, the Wellcome Trust and the School of Advanced Studies of the University of London.

collection development planning and service delivery. If successful, the outcomes could potentially be scaled up to provide an overview UK-wide.

These objectives are entirely complementary to RSLP's own vision, of facilitating the best possible arrangements for research support, in increasing the availability of information about the location of the UK's rich information resources.

### ***Proposal***

In response to RSLP's invitation, CURL is submitting a proposal to undertake a collection mapping analysis and survey. The proposal is for a small pilot project to evaluate the OCLC/Lacey iCas software as a useful methodology for analysing collections at an institutional level and also to provide a cross-institutional comparison in order to compare levels of overlap and uniqueness.

The records of three CURL and three non-CURL libraries will be analysed. The libraries participating represent different kinds of library: three are large university libraries with a broad range of subject coverage which reflects academic teaching and research in these universities; whilst the other three are more specialist libraries covering a smaller range of subjects but in much greater depth. The participants also have a range of different classification schemes: some use standard schemes, some in-house and some a combination of two or more systems. This reflects the reality of classification usage within CURL and more widely within the UK.

The project will also draw on the expertise of Dennis Nicholson as a consultant to the project. He is the Director of the Centre for Digital Library Research at the University of Strathclyde and is actively involved with a number of projects, notably HILT, SCONE and CAIRNS, which are all concerned with collection level descriptions, and the measurement of collection strengths as a means towards collaborative collection management.

### ***Libraries to be included***

The three CURL libraries are Edinburgh University Library, Imperial College London, which also includes the Science Museum Collection and Liverpool University Library. Together these libraries hold a total number of around 1.6 to 1.7 million electronic records.

The three non-CURL libraries are Hull University Library, the Natural History Museum and the School of Oriental and African Studies (SOAS) in London. Together these libraries hold a total number of 1.5 – 1.6 million records

Four of the libraries are participating in other collection analysis projects: Edinburgh is involved in the SCURL manual Conspectus project; Hull in RIDING and Imperial College and SOAS are involved in the M25 project.

### **Edinburgh University Library**

Edinburgh University Library (EUL) was founded in 1580 when Clement Litill, an Edinburgh Advocate, died, bequeathing his collection of 276 volumes to the Town and Kirk of Edinburgh. The library now contains a significant number of important collections, with around 660,000 bibliographic records on the CURL database. Edinburgh University Library holds collections in a wide range of subject areas to

support the university's teaching and research areas. Collections of particular note are: a collection of books and manuscripts of modern Scottish authors which is of international importance; English pre-1800 drama, literature of the 1930s, Carlyle and Scott. The library is decentralised and many of the departmental libraries also hold significant collections, including the Royal (Dick) School of Veterinary Medicine Library (founded 1825), New College Library (founded 1846) and the Reid Music Library (founded ca1860) and, most recently, Moray House Institute of Education Library (founded 1835).

### **Imperial College London and Science Museum Collection**

There are two main libraries at Imperial College: the Central Library which houses material on Computing, Life Sciences and Medicine, Management and Social Sciences, Science and Technology Studies and the Haldane Collection on Humanities. The Science Museum holds material which specialises in the History and Public Understanding of Science and Science Communication. Together the two libraries hold 720,000 items and there are 276,000 bibliographic records from Imperial College and the Science Museum on the CURL database. In addition a number of departmental libraries support specific teaching and research areas within the College.

### **Liverpool University Library**

Liverpool University Library has two main libraries, the Sydney Jones Library which holds material relating to the Arts, Social and Environmental Studies and Special Collections; and the Harold Cohen Library which has collections in Science, Medicine, Engineering and Veterinary Science. There are also a number of departmental libraries. The library's stock totals over 1.5 million volumes representing 688,000 bibliographic records, all of which are on the CURL database. Particular collection strengths include: Latin-American Studies, Irish Studies, Archaeology, English Literature, Mediaeval Studies, Medicine and Veterinary Science and Science generally.

### **Hull University Library**

The University of Hull has three libraries: the Brynmor Jones Library on the main campus; the Health Library on the East Riding campus and the Keith Donaldson Library on the Scarborough campus. These together house over 610,000 titles including 17,000 periodical titles, of which 3,500 are current. The collections support the learning, teaching and research activities of the University and the entire stock of over 1.2 million volumes is included on the library's Innopac catalogue. A number of the collections in the Brynmor Jones Library are of national significance and include material on Labour History, South-East Asian Studies, American Studies, 20<sup>th</sup> century English Literature and the European Resistance movement during World War II.

### **Natural History Museum**

The Library of The Natural History Museum contains the largest collection of natural history materials in the world. The 1 million volumes, include complete runs of early serial titles, of which some 23,000 titles are held and just under 10,000 are current.

Also held are around 500,000 original artworks, making this the third largest collection of art on paper in the UK. Other holdings include 10,000 manuscripts and 75,000 maps. Some 80% of the material has an electronic record on the Unicorn system which can be searched at <http://library.nhm.ac.uk/>

### **School of Oriental and African Studies (SOAS)**

The SOAS Library houses one of the world's finest collections of materials relating to the study of Asia, Africa and the Middle East. The collections cover a range of subjects in the social sciences and humanities and are arranged into seven regional (China, Japan, Africa, South East Asia, South Asia, Ancient Near East, Near and Middle East) and two subject collections. There are approximately 1.2 million volumes in the collection, and the Library receives around 4,500 current periodical titles.

### ***Deliverables***

The two main deliverables of the project will be the collection analysis results and a report looking at the successes of the project and any problems encountered during the analysis stage and how these might be resolved. It will also identify future possible developments.

### Analysis

There will be a collection analysis which will show the number of titles and the percentage level of the collections within each of the conspectus divisions, categories and subjects; and a title overlap report providing a summary of uniquely held and jointly held titles. The analysis can be customised by the individual libraries to some extent to take into account the different data held within the libraries.

Each library will receive an individual analysis to title level. CURL will receive a group analysis to the subject level, with overlap statistics and uniqueness measures. CURL will also receive a copy of the results for each individual library.

### Report

Once the analysis is received by CURL, a report will be produced demonstrating how successful the analysis was and identifying ways in which the software could be used more widely; it will also look in detail at any problems identified during the project.

Key points in the evaluation will include:

- Whether it is a practical solution in terms of cost and ability to deal with a range of database idiosyncrasies
- In what sense the method can be claimed to be truly objective.

The report will also make general recommendations on whether it is possible to compare the methodology with other collection level descriptions methodologies, and how this might be taken forward.

OCLC have also indicated that they are willing to provide £3,000 maximum to sponsor a working seminar to disseminate the outcomes of the project to the UK community, subject to the agreement by each party, OCLC, CURL and RSLP, that the project was successfully completed and that such a seminar would be valuable.

## ***Methodology***

### Preparation of data

Participants will complete the OCLC/Lacey ACAS Planning Guide and return it. For CURL libraries whose data are to be sent from CURL, much of this work will be done centrally by CURL. The planning guide will enable OCLC/Lacey to identify where classification information is held and how to identify other relevant data.

### Export of records to OCLC/Lacey

Records will be sent in one of two ways: either directly from libraries using ftp or from Manchester Computing using ftp. For the CURL libraries whose data are already held on the CURL database the records will be sent by Manchester Computing. Non-CURL libraries will send their records directly to ACAS at OCLC/Lacey.

### Analysis

OCLC/Lacey will de-duplicate records in order to find classmarks for as many records as possible. The records will then be matched against OCLC's WorldCat in order to find the remaining classmarks. The records can then be run against the conspectus software to analyse the collections. During this time there may be questions arising out of the data and library staff in the individual institutions may need to be contacted.

Once analysis is complete OCLC/Lacey will send the CD-ROM containing the analysis to CURL and to the participating libraries.

## ***Timetable***

### Month 1 Preparation

This will include collecting information from the libraries about which fields contain classification information and other information needed, and arranging the timetable for records to be sent to OCLC/Lacey.

### Month 2 – 4 Export

Records will be exported either by the participating libraries or by Manchester Computing on behalf of CURL

### Month 5 – 6 De-duplication, matching and analysis at OCLC/Lacey.

### Month 7 – 8 Report writing and dissemination.

CURL will provide a report to RSLP, evaluating the outcomes of the analysis, the usefulness of the methodology and will make recommendations for further work, if appropriate.

If considered useful by all parties, OCLC will host a seminar, as previously indicated.

## ***Project Costs***

The quotation from OCLC is at cost, and the usual charge of £8,000, to cover European overheads has been waived.

All costs are exclusive of VAT.

*Quotation for analysis of 3.2 m records	is \$81,000 @1.50	£54,000
Consultancy fees to Dennis Nicholson		£ 1,000
Staff time from CURL (data issues, report writing) calculated as an equivalent of 3 weeks FTE		£ 2,000

**Total** **£56,000**

Co-funding from RSLP	£28,000
Co-funding from CURL	£28,000

\*The price quoted from OCLC will not exceed \$81,000. Where it proves not possible to assign a class mark to a particular record, then no charge will be made. The actual cost is therefore likely to be less than quoted. This cannot be quantified at this stage.

### ***Other Partner Contributions***

Manchester Computing have agreed to undertake the data upload at no additional cost, but it will have an impact on other scheduled work for CURL/COPAC staff.

The non-CURL partners will have to provide staff time in assisting with data input. It is impossible to quantify this at this stage.

OCLC have agreed, in principle, to host a seminar if appropriate, following the outcomes of the analysis. This would be a maximum of £3,000. They are also willing to discuss areas of further research following the outcomes of the project.

#### **5.4 Letter of Agreement between CURL and OCLC (scanned from hard copy)**

OCLC Lacey Product Center  
4224 6<sup>th</sup> Ave.S.E., Bldg.3  
Lacey WA 98503-1 040 LISA

1 -800-342-5956  
(360) 923-4000  
Fax: (360) 923-4009  
www.oclc.org/western

April 10, 2001

Chris Bailey  
CURL Executive Secretary  
Consortium of University Research Libraries  
Glasgow University Library  
Hillhead Street  
Glasgow G12 8QE  
United Kingdom

Dear Ms. Bailey:

This Letter Agreement outlines the OCLC Automated Collection Analysis Services (ACAS®) proposed for the Consortium of University Research Libraries.

For each of six libraries designated by the Consortium of University Research Libraries, OCLC will perform individual automated collection analysis at the Conspectus division/category/subject/title level. For the six libraries as a group, OCLC will perform collection analysis and title overlap at the Conspectus division/category/subject level. Record processing will include matching against WorldCat to obtain call numbers for records lacking Library of Congress or Dewey Decimal System classification. OCLC's processing will use the WLN LC Conspectus structure and specifications defined by the libraries in the *ACAS Planning Guide (APG)*. The *APG* becomes part of the Letter Agreement. This agreement is valid from the date of signing by both parties through April 30, 2002.

The libraries participating in the test project include:

Edinburgh University Library  
University of Hull Library  
Imperial College of Science, Technology & Medicine, Central Library  
The University of Liverpool, Sydney Jones Library  
Natural History Museum  
School of Oriental and African Studies

#### **Automated Collection Analysis Reports (Age and Content)**

For the estimated 3,200,000 bibliographic records to be provided from their local systems by three individual libraries and from the CURL database for three of the libraries, OCLC will produce statistical reports showing the age and content of the individual libraries' collections, according to the WLN Conspectus structure at the division/category/subject/title level and parameters specified in the *APG*. The reports will consist of statistical tables and graph views of the collections' age and content based on the parameters specified in the *APG*. Each library's data will be output on an OCLC iCAS CD-ROM disc using the most current production version of the iCAS software. One set of reports for each library will be provided.

OCLC Online Computer Library Center, Inc.

A nonprofit, membership, computer library service and research organization  
Automated Collection Analysis Reports (Age, Content and Overlap)

OCLC will also produce statistical reports showing the age, content and overlap of the combined six libraries' collections, according to the WLN Conspectus structure at the division/category/subject level and parameters specified in the *APG*. The reports will consist of statistical tables and graph views of the collections' age, content and overlap based on the parameters specified in the *APG*. The data will be output on an OCLC iCAS CD-ROM disc using the most current production version of the iCAS software. Two sets of these combined resulted will be provided.

Based on the information provided by the Consortium of University Research Libraries in the *APG*, we have established the following processing schedule. Project schedules are based on the number of records estimated by the Consortium of University Research Libraries and the *APG* specifications submitted by the Consortium of University Research Libraries. Project modifications will require a re-evaluation by OCLC and possible changes to the schedule.

### Processing Schedule

Approximated file size 3,200,000	File Delivery Date by As soon as possible	Anticipated Completion Date Three months after receipt of readable data.
-------------------------------------	--	--

### Price Schedule

The following price schedule provides unit cost for age, content and overlap analysis. Final costs for the ACAS processing will be based on the actual number of bibliographic records submitted by the Libraries, the number of hours required for data processing and the options selected. The price proposal is included as APPENDIX A.

<b>ACAS Group Age, Content and Overlap Analysis</b>	
<i>Service</i>	<i>Unit Price</i>
Project management fee	\$450
Record intake and preparation	\$125 per hour per library, minimum 8 hrs per library
Base processing per library	\$500
Record processing - if total is:	
1-100,000 records	\$0.04 per record
100,001-250,000 records	\$0.035 per record
250,001-500,000 records	\$0.03 per record
500,001-750,000 records	\$0.025 per record
750,001-1,000,000 records	\$.02 per record
1,000,001-2,500,000 records	\$0.015per record
2,500,01-5,000,000 records	\$0.01 per record
5,000,001 or more records	\$.005 per record
Pricing above is for the basic age, content and overlap analysis at the WLN Conspectus division and category level. Additional options include:	
Division/category/subject reporting is:	
1-100,000 records	\$1000
100,001-200,000	\$2000
200,001-300,000	\$3000
300,001-400,000	\$4000
400,001 or more records	\$5000 per library
Display titles that comprise subject totals	\$500 per library plus \$0.005 per record
Resolve problems of nonstandard classification	\$125 per hour

Shipping and handling charges are included with the above processing charges.

**Project management:** \$450.00 is the initial fee for Automated Collection Analysis Services. This is a one time only fee when multiple analyses are completed.

**Record intake and preparation:** \$125.00 per hour per library covers intake of MARC format bibliographic records and holdings records via FTP from the consortium or library's local systems. It includes preparation of the bibliographic and holdings records for the analysis. Activities may include merging of record sources, UKMARC to USMARC conversions, and holdings conversions.

**Base processing:** \$500.00 per library covers basic analysis setup for each participating library.

**Record processing:** \$0.005 covers generating the analysis reports by matching the library's collection against the WLN Conspectus Table(s) at the division and category levels of the WLN Conspectus. Processing includes analysis of records rejected from the processing.

**Problem Resolution:** \$125.00 per hour covers problems of non-standard classification or classification errors of the records rejected from analysis and steps to include them in the results, including matching against WorldCat.

**File transfer and data output:** Standard OCLC EDX charges apply. Most OCLC libraries have a current EDX account that can be used to send and receive bibliographic and authority records. If the library already has an EDX account, the library need only complete an EDX assessment change form to add the ACAS file designation to its existing profile. Standard group charges are currently \$540.00 per year for unlimited file transfer of bibliographic records and authority records via FTP.

OCLC will invoice the Consortium of University Research Libraries when processing is complete and products are delivered. Unless otherwise agreed upon between the Consortium of University Research Libraries and OCLC, the Consortium shall make acceptance of OCLC services or products in writing within thirty (30) days from the date of receipt of such services or products. If the Consortium of University Research Libraries' written acceptance is not received by OCLC within thirty (30) days, services and products will be deemed accepted. The Consortium of University Research Libraries agrees to pay invoices thirty (30) days from the date of receipt of the invoice from OCLC. After the sixtieth (60) day from the date of receipt of invoice to the Consortium of University Research Libraries, interest will be charged on the unpaid balance at the rate of one and one-half percent per month. The Consortium of University Research Libraries agrees to pay such interest if incurred.

OCLC MAKES NO REPRESENTATIONS, WARRANTIES, OR GUARANTIES, EXPRESS OR IMPLIED, AND SPECIFICALLY DISCLAIMS ANY WARRANTY RELATED TO THE SERVICES, INCLUDING, BUT NOT LIMITED TO ANY IMPLIED WARRANTIES OR MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

EXCEPT AS PROVIDED BELOW, OCLC SHALL NOT BE LIABLE FOR ANY DAMAGES ARISING OUT OF OR IN CONNECTION WITH OCLC'S PERFORMANCE OF SERVICES UNDER THIS AGREEMENT, INCLUDING BUT NOT LIMITED TO INDIRECT, SPECIAL, INCIDENTAL, OR CONSEQUENTIAL DAMAGES AND LOSS OF ANTICIPATED REVENUE, PROFITS, GOODWILL, OR OTHER ECONOMIC LOSS. OCLC SHALL NOT BE RESPONSIBLE FOR UNINTENDED OR UNFORESEEN RESULTS OBTAINED BY the Consortium of University Research Libraries IN USING THE SERVICES.

ANY AND ALL DAMAGES THAT OCLC IS REQUIRED TO PAY, WHETHER FOR NEGLIGENCE, OMISSIONS, WILLFUL MISCONDUCT, BREACH OF CONTRACT, OR OTHERWISE, IN THE AGGREGATE, SHALL NOT EXCEED THE CHARGES ACTUALLY PAD BY the Consortium of University Research Libraries TO OCLC FOR THE SERVICES.

Notwithstanding any other provision of this Agreement, no party to the Agreement shall be deemed in default or breach of this Agreement for any delay, failure in performance, loss or damage due to any cause beyond its reasonable control.

I hope this Agreement meets with your approval. If so, would you please countersign in the space provided below and return one copy of this Letter Agreement to me.

Scott Barringer  
Director  
OCLC Lacey Product Center

Chris Bailey  
Executive Secretary  
Consortium of University Research Libraries

Date: 4/10/01

23/4/01

SB:ls

cc: Sally Loken  
ACAS Sr. Product Consultant

**APPENDIX A**

**iCAS PRICE PROPOSAL  
CONSORTIUM OF UNIVERSITY RESEARCH LIBRARIES  
10-Apr-01**

**Assumptions:**

Number of records in participating libraries	3,200,00 est.
Estimated record intake and preparation hours	72
Number of libraries	6

**Processing at the WLN Conspectus division/category level:**

	<u>Unit Price</u>	<u>Extended Price</u>
Project management	\$450.00	\$450.00
Record intake and preparation	\$125.00	\$9,000.00
UKMARC to USMARC conversion est. 40 hrs at \$125.00/hr	\$125.00	\$5,000.00
Matching records against WorldCat est. 100 hrs at \$125.00/hr	\$125.00	\$12,500.00
Base processing	\$500.00	\$3,000.00
Record processing \$125.00	\$0.005	16,000.00

**Additional Charges:**

Division/category/subject level reporting for 6 libraries at \$5000.00 each	\$5,000.00	\$30,000.00
Titles that comprise subject totals per library	\$500.00	\$3,000.00
per record	\$0.005	\$16,000.00
 Total:		 \$94,950.00
Less 15% discount:		\$14,242.50

**Final price will not exceed: \$80,707.50 \***

**Comments:**

Pricing valid through April 30, 3002.

\* Final pricing will be based on the actual number of bibliographic records analyzed and reported.

## **5.5 Guide to Using the Interactive Collection Analysis System**

### INTERACTIVE COLLECTION ANALYSIS SYSTEM

For

Consortium Of University Research Libraries (CURL)

December 2001

#### **INSTALLATION**

Your iCAS application will normally self-install. If you experience trouble during installation, copy the contents of the disc to a folder on your workstation or network and execute install.exe from there. For assistance, please call the OCLC Western Service Center Help Desk at 1-800-638-9956 between 6:30am and 5:30pm Pacific Coast Time.

#### **iCAS NAVIGATION**

After installation the iCAS software opens to a table view of the grand totals by publication date. To see other tables of the data, click on the "Select View" drop-down menu and select a table view. If the "Select Criteria" box appears, you are prompted to choose divisions and/or categories and/or subjects. Make your selections and click on SAVE. Then click on SHOW RESULTS.

To switch to a graph view of the data, click on the bar graph icon. A graph of the entire collection by publication date displays. To choose another graph, click on the "Select Graph" drop-down menu and select a graph. If the "Select Criteria" box appears, you are prompted to choose divisions and/or categories and/or subjects. Make your selections and click on SAVE. Then click on VIEW GRAPH.

To go back to the table views, click on the lightning bolt icon.

#### **FEATURES**

The iCAS CD-ROM provides collection analysis results in table and graph displays, using the OCLC/WLN LC Conspectus structure and the publication dates selected by the Library.

- The user chooses the level of detail for each view. The levels for your iCAS are the entire collection, Conspectus division, Conspectus category and Conspectus subject descriptor. These levels are arranged on the vertical axis in a hierarchical fashion.
- Publication date ranges are on the horizontal axis. The "Other" column contains the counts for records that have missing or invalid publication dates, and dates not covered by the specified date ranges.

- Report displays default to the highest level (the entire collection). Lower levels are “rolled up” and can be displayed, when selected, by clicking on the + in the left-hand column or by clicking on the number at the left of the top row.
- Pointing the mouse at a category or subject line provides, in the message bar at the bottom of the window, the complete name of the Conspectus category/subject and its call number range. The Conspectus line number of the category/subject is also displayed.
- When the mouse points to a numeric cell, a “tool tip” displays the cell’s percentage of the row.
- Under the “View” menu, the *Active Criteria* menu item shows the display criteria currently selected.

**Table views include:**

- Library Titles Held by Date (All): not searchable, summary of titles held by each library, subtotals for Audience optional
- Library Titles Held by Date (All Divisions, Select Library): not searchable, select single or multiple libraries, subtotals for Audience optional
- All Titles Held by Date (All Divisions): not searchable, summary of titles held in each Division, subtotals for Audience optional
- Library Titles Held by Date (Category Level, Select Div & Library): searchable on Category description, Conspectus Line number, Call number; select single or multiple libraries, select single or multiple divisions, subtotals for Audience optional
- Library Titles Held by Date (Subject Level, Select Div & Library): searchable on Subject and Category descriptions, Conspectus Line number, Call number; select single or multiple libraries, select single or multiple divisions, subtotals for Audience optional
- Library Titles Held by Date (Subject Level, Select Div, Cat & Library): searchable on Subject and Category descriptions, Conspectus Line number, Call number; select single or multiple libraries, select single division and single category, subtotals for Audience optional
- Titles Held by Division (Select Division and Library): not searchable, select single or multiple libraries, select single or multiple divisions, subtotals for Audience optional
- Titles Held by Category (Select Division and Library): searchable on category description, Conspectus Line number, Call number; select single or multiple libraries, single or multiple divisions, subtotals for Audience optional
- Titles Held by Subject (Select Division and Library): searchable on category and subject description, Conspectus Line number, Call number; select single or multiple libraries, single or multiple divisions, subtotals for Audience optional
- Titles Held by Subject (Select Div, Cat and Library): searchable on category and subject description, Conspectus Line number, Call number; select single or multiple libraries, single division and single category, subtotals for Audience optional
- Title Overlap
- Title Overlap (Division Level, select Division & Library)
- Title Overlap (Category Level, select Division & Library)

- Title Overlap (Subject Level, select Division & Library)
- Title Overlap (Subject Level, select Div., Cat. & Library)
- Uniqueness (All)
- Uniqueness (Division Level, select Division & Library)
- Uniqueness (Category Level, select Division & Library)
- Uniqueness (Subject Level, select Division & Library)
- Uniqueness (Subject Level, select Div., Cat. & Library)

### Graphs of the data:

- The vertical axis shows the number of titles *per year*. The date ranges represented by the bars may vary, so an average per year is calculated. The exceptions to this are titles in the “Pre-XXXX” and “Other” date ranges. The numbers for these are exact title totals; they are not averaged.
- The Conspectus level displayed depends on criteria selected by the user.
- Under the “Edit” menu, clicking on *Copy* places the graph on the clipboard.
- Under the “View” menu, the *Active Criteria* menu item shows the display criteria currently selected.
- The message bar displays check boxes to toggle the following options on and off, as is appropriate to the graph you are viewing:
  - Combine adult and juvenile
  - Show all points, do not scroll.
  - Omit non-specific dates from the display (Pre-XXXX and Other.)
  - Fix scale.
- **Graph views include:**
  - All Titles Held by Date
  - All Titles Held by Date (Select Division): Select single division
  - All Titles Held by Date (Select Category): Select single division and category
  - All Titles Held by Date (Select Subject): Select single division, category and subject
  - Library Titles Held by Date (Select Library): Select single or multiple libraries
  - Library Titles Held by Date (Select Library & Division): Select single or multiple libraries, select single division
  - Library Titles Held by Date (Select Library & Category): Select single or multiple libraries, select single division and category
  - Library Titles Held by Date (Select Library & Subject): Select single or multiple libraries, select single division, category and subject
  - All Titles Held by Division
  - All Titles Held by Category (Select Division): Select single division
  - All Titles Held by Subject (Select Division & Category): Select single division and category
  - Titles Held by Division (Select Library): Select single or multiple libraries

- Titles Held by Category (Select Division & Library): Select single or multiple libraries, select single division
- Titles Held by Subject (Select Division, Category & Library): Select single or multiple libraries, select single division and category
- Titles Held by Library (All)
- Titles Held by Library (Select Library): Select single or multiple libraries
- Uniqueness Summary (All)
- Uniqueness (Select Library)
- Uniqueness (Select Division & Library)
- Uniqueness (Select Division, Category & Library)
- Uniqueness (Select Div., Cat., Sub. & Library)
- Title Overlap (All Divisions, Select Library)
- Title Overlap (Select Division & Library)
- Title Overlap (Select Division, Category & Library)
- Title Overlap (Select Div., Cat., Sub. & Library)

## **SEARCH CAPABILITIES**

- The Search icon (binoculars) is available when your criteria include categories. When you do a search, you are searching the data as defined by the criteria you set.
- You may truncate searches for category names and Conspectus line numbers, e.g. BID0000.
- Call number searching must be exact (no truncation).
- A pop-up menu displays "Find Data" options when the user clicks on the Search icon on the tool bar or selects "Find Data" from the Tools drop-down menu. The options are:
  - Text to find
  - Text type (Category, Conspectus line, or Call Number)

## **PRINTING**

Printing tables:

- A pop-up menu displays print options when the user selects the Print icon on the tool bar, or selects "Print View" from the File menu, or presses Ctrl-P. All contents of the table view are printed. You may choose to print the percentages or not.  
You may choose to print selected date ranges.
- A print preview of the selected information appears after the system has processed the Print View request. The print preview pop-up box:
  - Allows the user to print the current page, the entire report or selected page ranges.
  - Allows the user to save the print preview to disk as a ".PRV" file.
  - Allows the user to retrieve a previously saved print preview, ".PRV" file.
  - Allows user to navigate to any page in the preview.

**Printouts:**

- Provide report titles, print date and analysis date at the bottom of each page.
- Can include cell percentages that were represented in the Table view by Tool Tips.

**Printing graphs:**

- To send a copy of the graph to the printer, select the Print Graph option from the File menu, or press Ctrl-P, or click on the printer icon.
- Headers for printed graphs include institution name, title of the graph, processing date for the underlying data and active criteria in effect.

**Export tables and graphs**

- A pop-up menu displays export options when the "Export View" or "Save Graph" item is selected from the File menu. A pop-up menu prompts the user for the export location and file name.
- All table displays can be exported to ".CSV" files and all graphs can be exported to ".JPG" files for manipulation by other software programs.
- A check box in the table view export dialog box provides the option to 'Include Header Row' in the exported file.

## **5.6 Project Summary by Sally Loken, OCLC ACAS Consultant**

### **CURL COLLECTION ANALYSIS PILOT PROJECT**

#### **PROJECT SUMMARY**

#### **OCLC AUTOMATED COLLECTION ASSESSMENT AND ANALYSIS SERVICES OCLC LACEY PRODUCT CENTER**

**DECEMBER 14, 2001**

On April 23, 2001 the Consortium of University Research Libraries (CURL) contracted with the OCLC Lacey Product Center (OLPC) through its Automated Collection Assessment and Analysis Services (ACAS), to analyze the machine-readable book format records of the following libraries:

Edinburgh University Library  
University of Hull Library  
Imperial College of Science, Technology and Medicine, Central Library  
The University of Liverpool, Sidney Jones Library  
Natural History Museum  
School of Oriental and African Studies

The analysis project included an individual collection analysis for each participating library, at the division/category/subject/title level of the WLN Conspectus, and a combined analysis at the division/category/subject level with overlap and uniqueness measures. CURL provided the publication dates and date ranges to be used for the age analysis.

The libraries sent 2,767,669 book format bibliographic records to OLPC. OLPC deduplicated the records in each library's file, arriving at a new total of 2,707,696 bibliographic records. Many of these records had no LC or Dewey classification numbers, which are required for OCLC content analysis. OLPC first matched the NO CALL NUMBER records of each library with the files of the other CURL libraries to obtain call numbers if present in the other libraries' records. Then OLPC sent the remaining NO CALL NUMBER records to the OCLC Office of Research for matching against WorldCat. These processes yielded call numbers for 570,705 records. During analysis processing 12,802 ill-formed call numbers were identified and rejected. The total number of records analyzed for the individual libraries is 2,268,225, which is 84% of the deduplicated total records sent. Only 16% of the records could not be processed for content analysis because of not having a call number. Results for each library are as follows:

University of Edinburgh:	604,531 - 78% of records were analyzed
University of Hull:	451,590 - 99.9% of records were analyzed
Imperial College:	202,758 - 74% of records were analyzed
University of Liverpool:	536,816 - 85% of records were analyzed
Natural History Museum:	45,820 - 47% of records were analyzed
SOAS:	426,710 - 92% of records were analyzed

The table below shows detailed information about the records for each library.

Library	Number of Records	Number of Books	Number of Class Fields Used	After WorldCat Matching	Call Numbers Added Using Other Libraries and WorldCat.
Edinburgh (EUX)	810,885	784,933	050: 326,386 082: 22,100 Total: 348,486 NOCALL: 436,447	Records: 772,923 Holdings: 606,124 Rejects: 1,593 On iCAS: 604,531 NOCALL: 166,799	269,648
Hull	473,074	473,069	050: 473,052 NOCALL: 17	Records: 460,653 Holdings: 460,647 Rejects: 9,057 On iCAS: 451,590 NOCALL: 6	11
Imperial College (LIP)	305,570	276,952	050: 61,793 082: 42,793 Total: 104,586 NOCALL: 172,366	Records: 274,398 Holdings: 203,137 Rejects: 379 On iCAS: 202,758 NOCALL: 71,261	101,105
Liverpool (LVT)	687,959	661,084	090: 432,171 NOCALL: 228,913	Records: 634,714 Holdings: 537,558 Rejects: 742 On iCAS: 536,816 NOCALL: 97,156	131,757
NHM (N8Q)	220,477	99,971	050: 1,963 090: 1,084 082: 695 092: 167 Total: 3,909 NOCALL: 96,062	Records: 97,590 Holdings: 45,873 Rejects: 53 On iCAS: 45,820 NOCALL: 51,717	44,345
SOAS (LOA)	496,828	471,660	082: 407,911 NOCALL: 63,749	Records: 467,418 Holdings: 427,688 Rejects: 978 On iCAS: 426,710 NOCALL: 39,730	24,019
	2,994,793	2,767,669			

The data in the column labeled **After WorldCat Matching** is useful in understanding the progression of processing results. For Edinburgh University Library, note that after WorldCat matching there were 772,923 book records, of which 606,124 had holdings (call numbers) and could be analyzed; there were 166,799 Edinburgh records with no call numbers at this point. The analysis processing rejected 1,593 call numbers that were ill-formed. Thus Edinburgh University's iCAS CD-ROM contains an analysis of 604,531 records.

Below is another way of looking at the data.

<b>Library</b>	<b>Books Sent</b>	<b>Total Records Left After Deduplication</b>	<b>Total Call Numbers After All Processing - Percentage of Total Records</b>	<b>Number of Call Numbers From OCLC Matching and Other CURL Project Libraries - Percentage of Total Call Numbers</b>	<b>Number of Records With No Call Numbers After Matching - Percentage of Total Records</b>
Edinburgh	784,933	772,923	606,124 - 78%	269,468 - 45%	166,799 - 22%
Hull	473,069	460,653	460,647 - 99.9%	11 - 0%	6 - 0%
Imperial	276,952	274,398	203,137 - 74%	101,105 - 50%	71,261 - 26%
Liverpool	661,084	634,714	537,558 - 85%	131,757 - 25%	97,156 - 15%
NHM	99,971	97,590	45,873 - 47%	44,345 - 97%	51,717 - 53%
SOAS	471,660	467,418	427,688 - 92%	24,019 - 6%	39,730 - 8%
<b>TOTALS</b>	<b>2,767,669</b>	<b>2,707,696</b>	<b>2,281,027 - 84%</b>	<b>570,705 - 25%</b>	<b>426,669 - 16%</b>

The combined, or group, analysis that OLPC prepared for CURL differs from the individual library analyses in three ways. The combined analysis does not include the individual library titles that comprise subject totals. It does include title overlap and uniqueness measures. And it also includes, at CURL's request, a brief analysis of the records with no call numbers. These records could not be analyzed for content, but we have analyzed them by age and holding library; overlap for these titles is also shown. To provide this information OLPC created a 25<sup>th</sup> Conspectus division, NO CALL NUMBER, which you will find listed alphabetically between MUSIC and PERFORMING ARTS in iCAS displays. Because these records are included in the combined analysis, the library totals are higher than they are in the individual library analyses.

On January 10, 2002, OLPC Product Manager Glenda Lins will deliver to CURL project participants the individual and group analysis results on iCAS CD-ROMs and will provide training on their use.

OLPC wishes to express our appreciation to CURL and the libraries for selecting OCLC to do this analysis work. We look forward to similar collaborations in the future.

Prepared by Sally Loken, OCLC ACAS Consultant

## 5.7 Additional Data from OCLC

Before Processing:

Library	Call Number Field(s)	Classification Scheme	Call Number Comments	Number of Records	Number of Class Fields	Number of Books	Bib Comments
Edinburgh (EUX)	050 and 082	LC and Dewey	some Dewey like \$a2.9 some 082s are blank some Dewey only have a \$2	810,897	339,097 and 166481 374953 with one 435932 with none	784,943	650 has indicator 4 Will also need matching to get call numbers for most records
Hull	050	LC	no decimal point	473,074	577,693 473,073 with 1 without	473,074	Needs UKMARC to MARC21 conversion 650 has no indicator
Imperial College (LIP)	050 and 082	LC and Dewey	999 field is not call numbers as APG says	305,570	63,187 and 91,661 106857 with one 198713 with none	276,952	
Liverpool (LVT)	090	LC		687,959	442,984 with 244,975 without	661,084	
NHM (N8Q)	NONE	Varied	No usable call numbers	220,477		99,971	
SOAS (LOA)	082	Dewey	look like Dewey with an alpha prefix	496,828	468960 with 27,868 without	471,660	
				2,952,208		2,725,087	

After Processing:

Library	Call Number Field(s)	Classification Scheme	Call Number Comments	Number of Records	Number of Books	Number of Class Fields Used	After WorldCat Matching	Call Numbers Added Using Other Libraries and WorldCat. Number of Missing Records
Edinburgh (EUX)	050 and 082	LC DDC	Did Biography routine. Changed Law to K1.	810,885	784,933	050: 326,386 082: 22,100 Total: 348,486 NOCALL: 436,447	Records: 772,923 Holdings: 606,124 Rejects: 1,593 On iCAS: 604,531 NOCALL: 166,799	269,648 12,010
Hull	050	LC		473,074	473,069	050: 473,052 NOCALL: 17	Records: 460,653 Holdings: 460,647 Rejects: 9,057 On iCAS: 451,590 NOCALL: 6	11 12,416
Imperial College (LIP)	050 and 082	LC DDC	999 field is not call numbers as APG says. Did Biography routine. Deleted leading X and Y. Changed Law to K1.	305,570	276,952	050: 61,793 082: 42,793 Total: 104,586 NOCALL: 172,366	Records: 274,398 Holdings: 203,137 Rejects: 379 On iCAS: 202,758 NOCALL: 71,261	101,105 2,554
Liverpool (LVT)	090	LC		687,959	661,084	090: 432,171 NOCALL: 228,913	Records: 634,714 Holdings: 537,558 Rejects: 742 On iCAS: 536,816 NOCALL: 97,156	131,757 26,370
NHM (N8Q)	050, 090, 082, 092	LC DDC		220,477	99,971	050: 1,963 090: 1,084 082: 695 092: 167 Total: 3,909 NOCALL: 96,062	Records: 97,590 Holdings: 45,873 Rejects: 53 On iCAS: 45,820 NOCALL: 51,717	44,345 2,381
SOAS (LOA)	082	DDC		496,828	471,660	082: 407,911 NOCALL: 63,749	Records: 467,418 Holdings: 427,688 Rejects: 978 On iCAS: 426,710 NOCALL: 39,730	24,019 4,242
				2,994,793	2,767,669		2,694,894/2,707,696	59,973 missing

Combined files and ran ripple to find call numbers.

2,767,669 Records in.

2,390,038 Records after deduping all records together.

377,631 Records deduped away.

997,554 Total NO CALL NUMBER holdings statements before deduping.

860,231 NO CALL NUMBER records sent to Ed.

436,544 Matched on WorldCat records with call number fields (050, 082, 090, 092).

2,390,083 Records into analysis

2,378,534 Records on iCAS

2,695,169 Holdings on iCAS

<b>Library</b>	<b>Books Sent</b>	<b>Total Records Left After Deduplication</b>	<b>Total Call Numbers After All Processing - Percentage of Total Records</b>	<b>Number of Call Numbers From OCLC Matching and Other CURL Project Libraries - Percentage of Total Call Numbers</b>	<b>Number of Records With No Call Numbers After Matching - Percentage of Total Records</b>
Edinburgh	784,933	772,923	606,124 - 78%	269,468 - 45%	166,799 - 22%
Hull	473,069	460,653	460,647 - 99.9%	11 - 0%	6 - 0%
Imperial	276,952	274,398	203,137 - 74%	101,105 - 50%	71,261 - 26%
Liverpool	661,084	634,714	537,558 - 85%	131,757 - 25%	97,156 - 15%
NHM	99,971	97,590	45,873 - 47%	44,345 - 97%	51,717 - 53%
SOAS	471,660	467,418	427,688 - 92%	24,019 - 6%	39,730 - 8%
TOTALS	2,767,669	2,707,696	2,281,027 - 84%	570,705 - 25%	426,669 - 16%

<b>Library</b>	<b># of call numbers found by matching each other</b>	<b># of call numbers found by matching WorldCat</b>
Edinburgh	59,385	210,263
Hull	5	6
Imperial College	29,984	71,121
Liverpool	36,439	95,318
Natural History Museum	8,800	35,545
SOAS	0	20,019

## 5.8 Responses from Partner Libraries

### From Edinburgh

EUL has had an opportunity to examine the OCLC iCAS CD-ROM data and the results appear to provide a fair analysis of Edinburgh University Library's collections. It does appear to do what it says, i.e. sorts our records by the LC classmarks found, fitting them in to the Conspectus categories and subjects. (Some mistakes, such as two 16th-century books on photography, are probably statistically insignificant, but they are a reminder that machine-matching is not 100 percent reliable.)

Two subject librarians with considerable experience of the EUL collections in their own area examined the data on the CD, as did the Library's database manager who has both knowledge of the collections themselves, and more importantly the data which has underpinned the analysis.

Colleagues looked at some specific subject areas on the CD, and, as far as possible, roughly compared the totals with the Library's catalogue, and the results seemed appropriate. We feel that it gives a reasonably accurate picture (in terms of Conspectus Categories) of EUL holdings but a number of concerns were raised and these are listed below:

- The results for the 'invented Conspectus sub-division' NO CALL NUMBER, does not appear on the CD, and as this figure is 22% (166,799) of the total (784,933) we would very much welcome the opportunity to have a look at these records.
- As with any type of Conspectus analysis, the subject groupings may not 'fit' the collections. There is always a difficulty of matching books with specific call numbers to another different subject listing. For internal organizational analysis, Edinburgh University Library may well want to compare holdings by site library, schools and/or colleges, in the future.
- The subject breakdown is fairly traditional and newer terms are not listed. Many of the categories out of date and/or confusing. A number of new ones are needed (e.g. Artificial Intelligence is not present under the Computer Science and Molecular Biology is not present under Biological Sciences) and some of the existing major categories should be renamed or amalgamated with others (e.g. Virology should be in Microbiology like Bacteriology). Some subject areas require major categories of their own - e.g. Veterinary Medicine is under Agriculture - surely this should be in a separate division?
- Although duplicates were excluded from the analysis, duplication of stock is an issue for consideration when planning for the future, in terms of space and funding.
- It would be useful for EUL if we could analyse by library site too.
- The numbers of titles in the 2000-2009 date range in some divisions is rather lower than expected.
- Some currently popular subject areas appear to be underrepresented – e.g. Remote Sensing only appears under Biological Sciences – Ecology and just

two titles are listed. EUL probably has over a hundred books on remote sensing and GIS.

- Other apparent anomalies include two books published between 1800 and 1899 placed in Special Topics in Computer Science – they are *Souvenirs de la vie militaire en Afrique* and *Lecons sur l'electricite et le magnetisme*.
- Subject coverage (in an individual library) should not be assessed on these figures alone, as any one title is classified only once; books on Japanese architecture may be found under Special types of building, for instance, rather than simply Architecture, Japan. This applies to the broadest categories (divisions) too, as for instance books on the Thatcher government may be in either Politics or History.
- Subject comparisons between libraries may be more logical, but only to the extent that the same titles are classified the same way in different libraries. This will presumably be increasingly the case, but it should not be assumed without evidence.
- We may have to be on our guard against departments asking for "their" books to be classified in "their" part of the classification, to improve the figures for use in TQA etc.

Richard Ovenden  
January 2002

---

#### From Hull

Chris,

We've reviewed Hull's data and it is accurate as far as analysis by LC classmarks is concerned. However, for some reasons lost in the mists of time we have not used the correct LC country subdivisions in some parts of the LC schedules. This occurs particularly in Art & Architecture, some parts of Geography and some parts of Business & Economics. So for example, instead of NA961-997 = Arts & Architecture. Architecture. Great Britain, we have NA961-997 = Arts & Architecture. Architecture. France & Germany.

The good news is that we can identify the areas where we have deviated from the standard and we can compile a list of our deviations.

Regards,

Bridget Towler  
31 January 2002

---

#### From Imperial (with replies from OCLC in bold)

##### Overlap / uniqueness method of study is flawed.

All representatives at the meeting in Birmingham were surprised at the level of uniqueness found when the 6 libraries were compared. We were told that titles were

being matched on author and title information. No use of date or ISBN appears to have been made. The reason given was that different editions of the same work could then be considered as an overlap. I think many of us at the meeting assumed that this meant that certain sub-field data of both the '100' and '245' fields were being compared. From looking at IC's data I do not think that is the case.

Data appears to have been extracted from the whole of the 245 field only. For both UK Marc and Marc21 records this will include the main title, subtitle and statement of responsibility, as they appear on the title page. For Marc21 it could also include the volume number and volume title that appear in the 248 field for UK Marc.

The examples I found within IC's data alone suggest that single character differences within the extracted string of the 245 field will result in 2 records not being recognised as essentially the same work. E.g. extract space, different punctuation, and the different appearance of the author's name.

Within my own data I found the following 2 items. They are different editions of the same work. It appears to be the presence of an extra space after the word 'students' and the abbreviation of 'part' that has caused these 2 records to be considered unique within IC's data.

Advanced mathematics for technical students by A. Greary, H.V., Lowery, and H.A. Hayden Pt.1.

Advanced mathematics for technical students by A. Greary, H.V., Lowery, and H.A. Hayden Part1.

OCLC need to describe in detail to CURL and RSLP how they are matching the titles to analyse overlap and uniqueness. On the basis of what I have conjectured above it is very unlikely that the full picture of overlap is being seen. Figures provided on overlap CD should therefore be used with caution.

**Actually for the overlap we did use more information as we always do for overlap. The specifications say to use Title, Author, Edition, Publisher, Date, Material Type, LCCN, ISBN, and ISSN. This does mean that if the forms in the 1xx are different, they won't be considered the same record. And if the 245\$abnp subfields are different, they will not be considered the same records.**

**We do not use the data in the \$c (statement of responsibility) subfield, unless the record doesn't have that information separately subfielded.**

**Extra spaces and punctuation difference will be ignored, but differences in abbreviation will not be ignored. So Pt 1 and Part 1 will be considered different items.**

Subject Classification appears to be very broad in places.

I asked our Maths librarian to look a bit closer at the use of subject classification. Although on the whole we thought the titles fell in the correct divisions, it seems that the classification achieved overall was often rather 'blunt'. For example when looking at examples of titles classified as 'numerical analysis' within our own collections, some were classified correctly under this subject heading whilst others appeared on the more general topic 'Mathematical analysis, general'. As we do not use either Dewey or LC classification at IC presumably the project relied on those supplied by

the publisher or possibly other CURL libraries. Unless the publisher specialises in a particular subject area it is likely that their classification for publication purpose will be rather broad.

A broad approach to subject classification in some areas will make the data analysis rather less valuable to the research community.

**We do rely on the call numbers supplied by the library, or by the records on WorldCat. Since we used both Dewey and LC call numbers, some generalization is inevitable because those classification schemes don't treat mathematics exactly the same.**

---

From Liverpool

We have done random checking of the data and haven't found anything unexpected. This is presumably because most of our stock is already classified by LofC. There is a slight disappointment that 15% of our titles could not be matched and my staff have commented that among the 15% lie probably some of our unique strengths. But I have found looking at the results extremely useful and feel that for Liverpool I would judge the pilot to have been a success.

.....

Frances M. Thomson  
6 February 2002

---

From the Natural History Museum (with replies from OCLC in bold)

Due to time limitations, I have only been able to look at 2 or 3 categories in each division and 1 to 3 subjects in each category. I have concentrated on those areas where the Museum would expect to have no or very few holdings, or where we would expect to be very strong.

The NHM supplied 220,477 monograph-format titles; 99,971 of those were accepted as "books" and 47% of those were analysed. The final breakdown therefore represents less than 21% of our monograph collection. **We used the MARC RecType/BibLevel in the Leader to determine "books", because "books" were the only material type included in this project. If we had known to ignore the RecType/BibLvl we would have processed everything.....just a communication misunderstanding.**

The project summary gives statistics of call numbers added from cross-matching with project libraries and WorldCat collectively. It would be interesting to have figures for how many matches were found within the project and how many in WorldCat, separately. **We could get this if needed. Most of the library's usable call numbers came from the matching we did with the other comparison libraries.**

Very few of the Museum's books carry Dewey or LC shelf-marks (those few where we have left them in imported records). The subject breakdown is therefore based on classifications applied to the same material in other libraries. This approach is clearly flawed because classification takes place in the context of a particular community's needs. The approach breaks down particularly when used to compare specialist and general collections. As WorldCat was used as the major reference point, many of the libraries used are north American. This is apparent in the inappropriate classification of non-US or international material as specifically concerned with America. **This is certainly a valid point, one which arises all the time when using one libraries call numbers to analyze someone else's collection. The "inappropriate classification" probably has more to do with the Dewey to LC correspondence, which doesn't get real specific.**

The Conspectus breakdown seems weak in biology, particularly in relation to animal and plant species. This has led to large numbers of our books being classified under Agriculture. There is also a weakness in relation to nature conservation, environmental management etc., which are classified Business & Economics. **The Conspectus does ride the fine line between too much detail and not enough. That's why we do custom tables for specialized libraries. Don't understand the last sentence--nature conservation looks like it's in Biology to me and environmental management can mean just about anything so it could be in Economics, Agriculture, Forestry, or Environmental Sciences.**

The breakdown generally falls into 3 camps:

- Titles that are "correctly" assigned (i.e. as we would have done it, manually): fewer than 50%.
- Classifications that are defensible but inappropriate in the context of our collection: the majority. **I'd just like to say here that you would probably find this to be true of any specialized library, since nobody agrees 100% with the Conspectus structure.**
- A substantial minority that are, more or less glaringly, wrong.

Below are some illustrative examples.

- Very many of the 371 titles under Biology, General are not general, e.g. "Fish biology", "On Mexican millipeds". **This could be from the Dewey to LC conversion which is pretty broad. We could also make the Biology Division even more specific than it already is.... Also there isn't a Subject "Biology, General" so these titles are those which didn't fall into one of the more specific subject call number ranges.**
- We would expect to find nothing or nearly nothing under Economics > Production, Industrial Management. The 59 titles placed there are wholly inappropriate. Most relate to environmental management, biodiversity etc. There is also an odd selection on assertiveness at work. **As in the example above there is no Subject "Economics - Production, Industrial Management" so these titles didn't fall within the two subject call number ranges. The book on assertiveness at work is about organizational behavior which falls within this call number range....Perhaps this is another area in the table where adding more subjects would be helpful.**
- 25 titles have been classified Education > Individual Institutions - United States because they were produced as thesis papers; their subject content has nothing to do with education. **This is where local practice comes in--some libraries classify theses under the call number for that institution, especially**

**colleges cataloging theses from their own institutions. These records probably matched in WorldCat.....**

- Law - United States, Federal contains 1 UK and many general, non-national texts. **These are probably from the Dewey call numbers, which only vaguely correspond to this category (there isn't an LC equivalent to Private Law, for instance).**
- Physical sciences > Science, General (108 entries) shows huge numbers of life science texts such as "Birds of Southeast Michigan". **Again, not really a Subject so this is a broad call number range. The "Birds of Southeast Michigan" title is classed by LC in Q11 (in this call number range).**
- 3 of 9 titles under Agriculture > Plant culture > Stone Fruit, .. are concerned with bananas. **The Subject is actually Stone Fruit, General, not Grape or Berry (not very descriptive, what it should say is Stone Fruit and other fruits not Grapes or Berries)**
- A number of titles under Decorative Arts, .. > Metalwork relate to diamonds or pearls. **LC has a broad category of class numbers under Metalwork, which includes diamond, pearls, even watches.**
- "Proceedings : Conference on the New Zealand Cretaceous" has been classified Vocational education. **Got us here. I'm thinking this was a product of the WorldCat match, which doesn't do real well with conference proceedings.....**
- Two books on fish - "Commercial fish of Australia", "Pescados y mariscos de las aguas Mexicanas" - are classified as Cookery - By Ingredient or Technique. **Both of these titles are about cooking or contain recipes.**
- "Clays and Shales of Michigan and their uses" is classified Organ music. **This record did have an LC call number M6, which is for Organ music.....there's no way to catch inaccurate classifications.**
- "Dominica, nature island of the Caribbean : a guide to geology, climate and habitats" appears under Water sports. **This looks like it matched against another in the series "Dominica, nature island of the Caribbean" which concerns diving in Dominica. Again this may be an example of the difficulties of matching with WorldCat.**
- 2 of the 3 records under Parapsychology are wholly inappropriate. **Can't explain this one, except that one of them has an inaccurate call number on WorldCat.**

In summary, we feel that this has been an interesting exercise but one with generally disappointing results. It has exposed the fact that automated approaches to the task need to rise several notches in sophistication before they can be of real practical use.

---

From SOAS (with replies from OCLC in bold)

Chris

I am slowly ploughing through. There are a lot of problems with SOAS data, but mainly caused by practice etc at our end. Much of the misposting - eg lots of Chinese history classed as Italian history are due to local classification arrangements - heaven knows why, but I'll get to the bottom of it. Another significant problem is best explained by illustration:

Our classmarks are prefixed by letters, followed by (allegedly) Dewey e.g. C 245.78 (C representing Chinese; we use up to three letters to reflect region, country and language).

In at least one case, the letter prefix is followed not by a Dewey number, but by an accession number: M 001, M 002 and so on. This explains why so many items were posted to generalities, library science etc. I haven't gone back to check, but I suspect that when Wallace Batchelor (now retired) filled in the OCLC form he forgot about these.

**OCLC response: Accession numbers are a big problem in a Dewey collection unless we are told explicitly and exclusively what to look for and ignore. Just telling us they use accession numbers isn't enough--how are we to tell what is and what isn't an accession number.**

I accept that these mis-postings are probably down to us, although it does cast interesting light on the OCLC methodology - having found a significant match between our records and World Cat, they seem then to have relied on our class marks. I know from Newcastle where we reclassified from around 6 strains of Dewey to the current one that different numbers can have significantly different meanings in different editions.

**OCLC response: They seem surprised that we would use their class marks over WorldCat's. But that was the plan all along--we only checked WorldCat for those records which didn't have a call number.**

Keith Webster

**5.9 Examples of Statistical Tables - Uniqueness (Division Level, select Division & Library) from combined analysis CD-ROM**

INSTITUTION	DIVISION	Unique	Shared By 2	Shared By 3	Shared By 4	Shared By 5	Shared By 6
Edinburgh University	AGRICULTURE	10144	1825	411	82	11	0
Edinburgh University	ANTHROPOLOGY	3228	1510	489	102	10	0
Edinburgh University	ART AND ARCHITECTURE	19706	4278	1002	134	10	0
Edinburgh University	BIOLOGICAL SCIENCES	10092	4611	1795	567	59	1
Edinburgh University	BUSINESS AND ECONOMICS	34319	13252	5806	1394	89	0
Edinburgh University	CHEMISTRY	2284	1209	464	127	4	0
Edinburgh University	COMPUTER SCIENCE	5969	2716	716	182	1	0
Edinburgh University	EDUCATION	11924	4126	539	62	2	0
Edinburgh University	ENGINEERING AND TECHNOLOGY	11646	4605	1386	338	7	0
Edinburgh University	GEOGRAPHY AND EARTH SCIENCES	6868	3081	1394	439	69	4
Edinburgh University	HISTORY AND AUXILIARY SCIENCES	55244	15600	4731	707	34	2
Edinburgh University	LANGUAGE, LINGUISTICS, AND LITERATURE	89233	24058	7545	648	24	0
Edinburgh University	LAW	12591	2737	943	110	6	0
Edinburgh University	LIBRARY SCIENCE, GENERALITIES AND REFERENCE	8760	2621	675	121	10	0
Edinburgh University	MATHEMATICS	6970	3513	1631	550	5	0
Edinburgh University	MEDICINE	25203	9023	2189	363	7	1
Edinburgh University	MUSIC	5774	1666	697	159	2	0
Edinburgh University	NO CALL NUMBER PRESENT	164711	2025	63	0	0	0
Edinburgh University	PERFORMING ARTS	2937	953	301	28	0	0
Edinburgh University	PHILOSOPHY AND RELIGION	51817	9790	2366	286	9	0
Edinburgh University	PHYSICAL EDUCATION AND RECREATION	3398	690	175	8	0	0
Edinburgh University	PHYSICAL SCIENCES	7899	3623	1353	402	6	0
Edinburgh University	POLITICAL SCIENCE	13014	5311	1940	333	18	0
Edinburgh University	PSYCHOLOGY	5412	1948	723	106	4	0
Edinburgh University	SOCIOLOGY	17519	8727	3528	632	53	0

Imperial College	AGRICULTURE	1904	826	300	72	11	0
Imperial College	ANTHROPOLOGY	315	190	105	32	10	0
Imperial College	ART AND ARCHITECTURE	2683	961	401	119	9	0
Imperial College	BIOLOGICAL SCIENCES	7850	4268	1694	555	59	1
Imperial College	BUSINESS AND ECONOMICS	10278	4513	2444	1072	89	0
Imperial College	CHEMISTRY	5077	1722	562	132	4	0
Imperial College	COMPUTER SCIENCE	5524	2929	748	182	1	0
Imperial College	EDUCATION	868	490	295	55	2	0
Imperial College	ENGINEERING AND TECHNOLOGY	32111	6297	1538	338	7	0
Imperial College	GEOGRAPHY AND EARTH SCIENCES	7068	3077	1430	440	66	4
Imperial College	HISTORY AND AUXILIARY SCIENCES	4281	1808	915	420	31	2
Imperial College	LANGUAGE, LINGUISTICS, AND LITERATURE	8245	2244	1054	417	24	0
Imperial College	LAW	698	265	144	51	7	0
Imperial College	LIBRARY SCIENCE, GENERALITIES AND REFERENCE	2311	759	315	105	10	0
Imperial College	MATHEMATICS	6129	3302	1520	549	5	0
Imperial College	MEDICINE	14322	5796	1662	351	7	1
Imperial College	MUSIC	1207	664	404	154	2	0
Imperial College	NO CALL NUMBER PRESENT	69926	1278	57	0	0	0
Imperial College	PERFORMING ARTS	487	155	68	23	0	0
Imperial College	PHILOSOPHY AND RELIGION	1537	834	404	160	9	0
Imperial College	PHYSICAL EDUCATION AND RECREATION	589	83	23	6	0	0
Imperial College	PHYSICAL SCIENCES	11782	4159	1412	398	6	0
Imperial College	POLITICAL SCIENCE	1526	719	374	197	18	0
Imperial College	PSYCHOLOGY	810	442	226	86	4	0
Imperial College	SOCIOLOGY	1884	1194	773	409	52	0
Natural History Museum	AGRICULTURE	2709	409	105	24	7	0
Natural History Museum	ANTHROPOLOGY	390	65	25	8	2	0
Natural History Museum	ART AND ARCHITECTURE	539	123	50	9	6	0
Natural History Museum	BIOLOGICAL SCIENCES	19214	2410	798	265	57	1
Natural History Museum	BUSINESS AND ECONOMICS	243	81	35	15	4	0
Natural History Museum	CHEMISTRY	191	66	28	12	4	0

Natural History Museum	COMPUTER SCIENCE	24	16	9	4	1	0
Natural History Museum	EDUCATION	49	8	2	0	0	0
Natural History Museum	ENGINEERING AND TECHNOLOGY	1210	201	82	30	3	0
Natural History Museum	GEOGRAPHY AND EARTH SCIENCES	7671	1256	633	245	53	4
Natural History Museum	HISTORY AND AUXILIARY SCIENCES	1357	247	91	23	7	2
Natural History Museum	LANGUAGE, LINGUISTICS, AND LITERATURE	200	47	18	8	2	0
Natural History Museum	LAW	95	23	10	3	3	0
Natural History Museum	LIBRARY SCIENCE, GENERALITIES AND						
Natural History Museum	REFERENCE	858	163	64	35	9	0
Natural History Museum	MATHEMATICS	37	13	14	9	3	0
Natural History Museum	MEDICINE	732	167	101	25	3	1
Natural History Museum	MUSIC	1	0	0	0	0	0
Natural History Museum	NO CALL NUMBER PRESENT	51234	460	23	0	0	0
Natural History Museum	PERFORMING ARTS	7	0	0	0	0	0
Natural History Museum	PHILOSOPHY AND RELIGION	127	39	10	7	0	0
Natural History Museum	PHYSICAL EDUCATION AND RECREATION	185	12	1	1	0	0
Natural History Museum	PHYSICAL SCIENCES	989	198	95	34	3	0
Natural History Museum	POLITICAL SCIENCE	40	12	8	3	0	0
Natural History Museum	PSYCHOLOGY	27	11	8	5	1	0
Natural History Museum	SOCIOLOGY	123	57	34	16	3	0
School of Oriental and African Studies	AGRICULTURE	2497	314	75	25	7	0
School of Oriental and African Studies	ANTHROPOLOGY	10046	1131	312	80	8	0
School of Oriental and African Studies	ART AND ARCHITECTURE	13781	919	141	21	8	0
School of Oriental and African Studies	BIOLOGICAL SCIENCES	1803	230	65	21	4	1
School of Oriental and African Studies	BUSINESS AND ECONOMICS	38318	4524	1442	505	86	0
School of Oriental and African Studies	CHEMISTRY	117	3	1	0	0	0
School of Oriental and African Studies	COMPUTER SCIENCE	122	14	16	8	0	0

School of Oriental and African Studies	EDUCATION	3704	414	121	21	2	0
School of Oriental and African Studies	ENGINEERING AND TECHNOLOGY	2424	181	52	30	4	0
School of Oriental and African Studies	GEOGRAPHY AND EARTH SCIENCES	16365	743	217	81	29	4
School of Oriental and African Studies	HISTORY AND AUXILIARY SCIENCES	55973	6783	1681	383	31	2
School of Oriental and African Studies	LANGUAGE, LINGUISTICS, AND LITERATURE	91148	4506	1047	281	22	0
School of Oriental and African Studies	LAW	10857	587	177	60	5	0
School of Oriental and African Studies	LIBRARY SCIENCE, GENERALITIES AND REFERENCE	10834	771	118	32	1	0
School of Oriental and African Studies	MATHEMATICS	413	44	24	8	2	0
School of Oriental and African Studies	MEDICINE	1791	212	80	26	4	1
School of Oriental and African Studies	MUSIC	2468	171	40	5	2	0
School of Oriental and African Studies	NO CALL NUMBER PRESENT	39565	161	4	0	0	0
School of Oriental and African Studies	PERFORMING ARTS	1481	158	42	11	0	0
School of Oriental and African Studies	PHILOSOPHY AND RELIGION	56069	3300	591	156	9	0
School of Oriental and African Studies	PHYSICAL EDUCATION AND RECREATION	493	26	6	1	0	0
School of Oriental and African Studies	PHYSICAL SCIENCES	1471	122	45	13	3	0
School of Oriental and African Studies	POLITICAL SCIENCE	36687	2346	602	172	18	0
School of Oriental and African Studies	PSYCHOLOGY	3637	168	53	23	3	0
School of Oriental and African Studies	SOCIOLOGY	24045	2297	707	288	51	0

University of Hull	AGRICULTURE	3033	485	199	65	10	0
University of Hull	ANTHROPOLOGY	3321	753	395	94	10	0
University of Hull	ART AND ARCHITECTURE	9930	1855	725	126	9	0
University of Hull	BIOLOGICAL SCIENCES	6741	1701	937	465	58	1
University of Hull	BUSINESS AND ECONOMICS	44611	9902	4695	1312	89	0
University of Hull	CHEMISTRY	2654	902	431	131	4	0
University of Hull	COMPUTER SCIENCE	1936	741	501	179	1	0
University of Hull	EDUCATION	1673	312	205	50	2	0
University of Hull	ENGINEERING AND TECHNOLOGY	8010	1752	797	318	7	0
University of Hull	GEOGRAPHY AND EARTH SCIENCES	10432	1690	804	345	66	4
University of Hull	HISTORY AND AUXILIARY SCIENCES	43181	9786	3935	665	34	2
University of Hull	LANGUAGE, LINGUISTICS, AND LITERATURE	96072	18527	6888	631	24	0
University of Hull	LAW	7088	1863	856	108	7	0
University of Hull	LIBRARY SCIENCE, GENERALITIES AND REFERENCE	5625	964	535	110	10	0
University of Hull	MATHEMATICS	4813	1989	1211	541	5	0
University of Hull	MEDICINE	7346	2274	1236	343	7	1
University of Hull	MUSIC	5761	917	542	159	2	0
University of Hull	NO CALL NUMBER PRESENT	6	0	0	0	0	0
University of Hull	PERFORMING ARTS	3474	737	275	27	0	0
University of Hull	PHILOSOPHY AND RELIGION	21993	5528	2046	268	9	0
University of Hull	PHYSICAL EDUCATION AND RECREATION	1270	403	158	8	0	0
University of Hull	PHYSICAL SCIENCES	5688	1435	799	381	6	0
University of Hull	POLITICAL SCIENCE	15493	3856	1735	327	18	0
University of Hull	PSYCHOLOGY	3927	1375	617	101	4	0
University of Hull	SOCIOLOGY	20815	5713	3095	609	53	0
University of Liverpool	AGRICULTURE	4920	1123	275	76	9	0
University of Liverpool	ANTHROPOLOGY	1686	597	267	92	10	0
University of Liverpool	ART AND ARCHITECTURE	13182	2958	861	131	8	0
University of Liverpool	BIOLOGICAL SCIENCES	10430	2848	1356	499	58	1
University of Liverpool	BUSINESS AND ECONOMICS	39727	10780	5042	1334	88	0
University of Liverpool	CHEMISTRY	2677	1056	440	126	4	0

University of Liverpool	COMPUTER SCIENCE	2682	742	476	177	1	0
University of Liverpool	EDUCATION	11446	3664	497	60	2	0
University of Liverpool	ENGINEERING AND TECHNOLOGY	14344	3366	1254	318	7	0
University of Liverpool	GEOGRAPHY AND EARTH SCIENCES	8564	2231	1015	374	62	4
University of Liverpool	HISTORY AND AUXILIARY SCIENCES	62659	12198	4094	674	33	2
University of Liverpool	LANGUAGE, LINGUISTICS, AND LITERATURE	105849	21292	6971	631	24	0
University of Liverpool	LAW	6170	1989	828	108	7	0
University of Liverpool	LIBRARY SCIENCE, GENERALITIES AND REFERENCE	10634	2026	618	117	10	0
University of Liverpool	MATHEMATICS	7244	2843	1522	543	5	0
University of Liverpool	MEDICINE	22396	4958	1788	352	7	1
University of Liverpool	MUSIC	2927	1118	564	159	2	0
University of Liverpool	NO CALL NUMBER PRESENT	95958	1150	48	0	0	0
University of Liverpool	PERFORMING ARTS	1804	573	241	27	0	0
University of Liverpool	PHILOSOPHY AND RELIGION	20720	5849	2020	267	9	0
University of Liverpool	PHYSICAL EDUCATION AND RECREATION	1634	484	168	8	0	0
University of Liverpool	PHYSICAL SCIENCES	7132	2441	1258	388	6	0
University of Liverpool	POLITICAL SCIENCE	12440	3956	1701	328	18	0
University of Liverpool	PSYCHOLOGY	4319	1716	635	103	4	0
University of Liverpool	SOCIOLOGY	19827	5784	3044	598	53	0

## **5.10 What we learned – Suggested Improvements for Future Analysis, from OCLC**

### **CURL COLLECTION ANALYSIS PILOT PROJECT**

#### **WHAT WE LEARNED – Suggested Improvements for Future Analyses**

**March 3, 2002**

On January 10, 2002, the results of the Consortium of University Research Libraries (CURL) pilot project were released to all the participating libraries through a training session conducted by OCLC Lacey Product Center staff. Also provided to participating libraries was a summary of results of the analysis and step-by-step description of the approach used to complete project.

The participating libraries and representatives from CURL had the opportunity to review their individual library's data and the summary data for the group, which provided overlap and uniqueness information. Because of time limitations, libraries were asked to further analyze their data and provide responses to CURL by the end of January.

Summaries were provided by all six participating libraries, reporting both positive and negative responses to the methodology used for the analysis, structure of the Conspectus and local classification issues. Some of the responses/issues are addressed below:

#### ***Methodology***

The methodology used was a first for OCLC Lacey Product Center. Most of libraries using the iCAS product classify their materials using either LC or Dewey Classification schemes. Since many of the CURL libraries – especially the pilot libraries – use schemes that are neither LC nor Dewey or use no classification scheme, we were provided an opportunity to try other ways to gather call numbers for the libraries records. The methodology was appropriate for most of the libraries since we were able to find call numbers for 84% of the libraries' records.

When libraries lack call numbers in their records and they have unique titles, these titles cannot be reported. This was a concern of a couple libraries that their unique materials did not appear in the results. The methodology used to match materials with no call numbers to OCLC WorldCat, did however, provide for the inclusion of more records than expected.

Because LC and Dewey Classification schemes do not treat all subjects the same way, there was some disagreement from the libraries about where their materials were mapped in the Conspectus. Since we matched the libraries records (titles, authors, LCCN's, Edition, Publisher, Date, Material Type, ISBN and ISSN) with OCLC WorldCat, we selected whichever call number available in the matching record, whether it was a Dewey or LC call number. Once again, this type of matching allowed for the inclusion of many extra records.

One of the libraries did in fact find several incorrect mappings to the Conspectus. We appreciate this and have already made the necessary changes to the Conspectus structure and these problems are eliminated for future analyses. Another library indicated the Conspectus uses “fairly traditional” terms and should be revised to include some of the newer terms.

### ***Lessons Learned***

As with all projects involving multiple libraries and participants, there were some misunderstandings between the participants. Some of the comments addressed areas of concern such as ‘it would have been nice to get the overlap information for all the libraries’ or ‘it would have been helpful if I had the No Call Number Conspectus division on my individual analysis’. This is mentioned to avoid the same type of confusion next time.

Also, understanding the methodology OCLC uses and the profiles created by the requesting library is important for the participants. One of the libraries sent us records that included books and monographic serials, but because of the specifications of the project, the monographic serials were not included in the analysis. This way, the expectations are the same for each participating library.

### ***Summary***

The limitations encountered from the lack of call numbers in data received from the CURL libraries required OCLC to investigate other opportunity to explore new methods for finding usable call numbers in order for us to complete the analysis. Though the methods used by OCLC did not include all the records from the comparison libraries, it did include the majority. One of the unfortunate consequences of not having usable call numbers in the records is the fact that the most unique holdings are not represented in the results.

With an 84% analysis rate, the CURL pilot project is deemed successful. OCLC hopes to further our relationship with UK libraries using iCAS. With a better understanding of the specifications and methodology, the next analyses will be even more successful.

**5.11 External Evaluators Report by Dennis Nicholson, Director of the Centre for Digital Library Research, University of Strathclyde**

**CURL Study of the OCLC/Lacey iCAS Software**

**1. The Context of the Study: Collaboration is now a Necessity**

***Co-operation in the provision of information resources and associated finding aids has become essential as well as desirable. Increasingly, the 'collections' required to meet the needs of researchers are geographically distributed - regionally, nationally, even globally. Accordingly, information managers aiming to meet these needs must themselves co-operate beyond institutional boundaries. Distributed networked collections need collaborative management if they are to meet the needs of users in a cost-effective fashion, and they need collaborative standards-based description if they are to be findable as and when required. Coherent distributed virtual 'libraries' won't just happen – practitioners must co-operate to manage the distributed collections and the associated user retrieval environments. To do so they need reliable standards-based tools that facilitate the objective analysis of both institutional and inter-institutional collections, and that do so, for preference, through automated methods that minimise the commitment of expensive staff time that might otherwise be required to support such an exercise. The iCAS tool evaluated by CURL in this study offers an automated method of:***

- ***Analysing institutional collections down to item level and mapping the results - via DDC or LC classifications – to the Conspectus subject headings used world-wide to measure and encode collection strength (Used in Australia, the US, and Scotland, for example)***
- Providing a detailed subject breakdown of institutional collections over time
- Identifying areas of inter-institutional overlap or uniqueness

The study results not only illuminate both the positive features of this method of collections assessment, and its potential and current limitations, but also:

- Show how results might be improved through future action on metadata interoperability and other areas,
- Highlight the importance of other ongoing UK work in the field
- Identify areas where additional research is needed to take the collaborative collection development agenda forward.

As such, they should help inform the development of a UK-wide collection mapping strategy, compatible (as it must be) with international work in the area.

**2. Automated Collection Mapping: The Requirement**

Collection mapping data is potentially valuable as management information to support:

- Efficient institutional, inter-institutional, regional, national, and international collection management policies and practices
- The provision of effective and reliable guidance for users and potential users of both institutional and distributed collections

In Scotland, for example, data from a SCURL inspired Conspectus based exercise carried out in the mid 1980s in the eight older universities and the two main public libraries in Edinburgh and Glasgow, and updated and added to on an ongoing basis since then, is used to:

- Guide users either to strong individual teaching, research or comprehensive collections in particular subject areas under the Conspectus scheme and link them to the appropriate online catalogues to search for specific items of interest, or to groups of geographically distributed subject collections that can then be cross-searched via Z39.50
- Support collaborative collection management and development by providing the same facilities for staff to assist them in recording and identifying collaborative collecting responsibilities, gaps in Scotland-wide or regional collections, and potential collaborative partners for closer co-operation in particular areas

One SCURL concern being addressed by the RSLP-funded SCONE project<sup>2</sup> is that in the past assessments in SCURL libraries have been regarded as too subjective and labour intensive. The project has accordingly been asked to examine possible ways of improving the objectivity of the collection strength measurement process, whilst at the same time reducing the effort and time involved in carrying it out. An examination of the iCAS approach investigated by CURL was outwith SCONE resources, and an attempt to obtain funding failed. SCONE accordingly concentrated on manual methods, and on a theoretical analysis of the strengths and weaknesses of an automated approach, in the knowledge that the costs of an automated analysis would have been regarded by many SCURL libraries as unattractive given other pressures on the local purse. SCONE, which reports in May 2002, has concluded that objectivity and low labour intensity is best achieved in SCURL libraries through:

- A short to medium term programme based on informed professional judgement. This to be constrained – in the interests of objectivity - by informal peer review and a functional focus on user navigational and discovery needs and staff local and collaborative collection development needs. Online interaction, mediated through the SCAMP collection management portal, would mediate this process.
- A linked gradual progression towards an increasingly automated approach in the longer term, this being regarded – potentially at least - as a key element in any future approach to reducing effort, improving objectivity, and enhancing staff tools for collaborative collection development on the one hand user facilities for accessing strong collections on the other.

The recognition of the potential importance of the automated approach reflected in this latter point is a direct result of examining the work of the CURL iCAS study in the

---

<sup>2</sup> See <http://scone.strath.ac.uk/>

light of a SCONE analysis of the strengths and weaknesses of an automated approach.

### *3. The Potential and Problems of an Automated Approach*

The outcome of this examination is presented in tabular form in Appendix A. This charts:

- 3.1** The optimum position possible with regard to the extraction of collection strength data as determined through the SCONE analysis
- 3.2** The current position in respect of the capabilities of the iCAS software and the metadata situation in the UK, as evidenced by the CURL study libraries and the libraries in CAIRNS
- 3.3** The implications of a comparison of the two
- 3.4** Implied actions required to improve the situation ('Solution Paths')
- 3.5** Additional relevant comments

The results of the analysis charted in the table support the following conclusions:

- a. Automated methods offer the potential of providing a wholly objective means of measuring most aspects of the idea of collection strength from item level metadata, although:
  - i. Some aspects such as the 'significance' of a collection may always entail the need for human assessment may always entail the need for human assessment
  - ii. The success of any automated method is dependent on the quality of the item level metadata available, the percentage of the collection that is catalogued, the data elements recorded in each item level record, the relevance of these data elements to user and staff needs, the compatibility of the metadata with the automated method in question (e.g. whether or not an iCAS compatible classification scheme is employed), and a number of other factors.
  - iii. Objectivity depends not just on objective methods of measurement, but on objective description, and an understanding of the limits of what may be inferred about the collection from the element measured (e.g. a numerically large research collections may be a 'strong' collection to a researcher but a 'weak' collection to a learner)
- b. Assessing the value of automated methods in general, and of the iCAS approach in particular, requires an examination of a number of factors not associated with the efficacy of the approach in respect of objective measurement. These are mainly associated with costs and needs and include:
  - i. The cost to the institution or group of institutions of employing the method concerned
  - ii. Whether the cost is within the means of the institution or group concerned
  - iii. The extent to which the requirement in terms of the needs of users and collection managers is understood in detail
  - iv. The extent to which the requirement is met by the method in question
  - v. The extent to which benefits are seen to justify costs, regardless of whether the costs are within the means of the institution or group

Very little hard and reliable information is available on these various factors. However, it is likely that there would be a split in UK libraries between those willing and able to fund an iCAS approach and those opting for a non-automated approach such as is proposed by SCONE in the short term.

- c. The iCAS approach is of potential value now to libraries whose collection is entirely covered by its online catalogue, or where collections not covered are readily susceptible to human assessment. However, all of the various points noted at (a) and (b) above are also a consideration, and may in many cases be a major one. Its value to groups such as CURL depends to a large degree on its value to individual CURL member institutions, although the fact that a group like CURL has a co-operative focus presumably tends to mean that the importance of measuring collection strength is itself increased for CURL member institutions.
- d. It is possible in the longer term to envisage collection strength measurement being carried out automatically by local library management systems at marginal cost to institutions. This would require:
  - i. Good, complete, cross-compatible metadata that entailed elements such as class number, subject, research or teaching code, language, location, access category, unique item identifier
  - ii. A collection strength index incorporating these elements at each of a number of institutional catalogues or, in cases of systems like COPAC, at a central site
  - iii. Interface mechanisms to allow staff and users to search or cross-search the indices and interact intelligently with the data in the indices

This approach would not only be low-cost (assuming good metadata) but would provide largely objective, up to the minute information, in contrast to the (current) iCAS approach which is (perforce) based on relatively infrequent snapshots. Strength in this scenario would be measured by numerical strength within categories such as research or teaching. Collections with low numerical strength but containing unique items would be evident, although elements such as the 'significance' of a collection would require subjective professional judgement.

- e. Further research is required with a view to determining:
  - i. The exact requirements of funding bodies, institutions, inter-institutional groups, collection managers, and users in respect of collection strength measurements
  - ii. Which elements of collection strength can be objectively measured and which must always entail subjective judgement
  - iii. How these may best be measured and any limitations on what we may legitimately infer from them as regards the strength of the collection in question
  - iv. Whether, given the actual requirements identified through e(i) above, there is a significant difference between objectively measured strength assessments and those based on (subjective) professional judgement
  - v. Comparative costs of at least four approaches to measuring collection strength (iCAS, the library management based system suggested in this report, Conspectus, and the proposed SCONE alternative), and an assessment of costs against benefits of each in a variety of circumstances (e.g. for CURL, for SCURL, and for the DNER, where legacy metadata problems are probably more manageable).

- f. The outcome of this research may show that there is one best short term method for all and one best long term aim. More likely, however, it will show that, in the short to medium term, different approaches are sensible in different circumstances - the SCONE approach, iCAS, or, possibly, where problems with legacy metadata are minimal, the collection index based approach suggested at (d) above.
- g. Problems with legacy metadata are a barrier to advancement in this area, as in so many others.

#### **4. Summary and Recommendations**

There is value in pursuing the goal of an approach to collection strength measurement that is either wholly or largely automated. Some institutions (such as Hull University in the present study) may already be in a position to take some advantage of it through iCAS, and this may also apply to some inter-institutional groups, but for most institutions and groups in the UK, there is much work to be done before the potential of an automated approach can be fully realised. Taking this work forward requires research work in various areas, undertaken with a view to informing the development of an agreed collection mapping strategy and collaborative collecting programme. This should be co-ordinated at UK-wide level but should probably be based – at a sectoral, regional cross-sectoral, or special interest (e.g. research interest) level – on smaller groups with a track record in the area such as CURL and SCURL. It is important that it be based on consensus and be as inclusive as possible, and that it therefore recognise the weaknesses of the automated approach in the current situation, whilst at the same time, aiming to make it the future basis of the strategy.

The automated approach has great potential, offering the long-term possibility of objective measurement, marginal costs, support for deep resource sharing, immediacy of collection 'strength' data, and a much improved service to users, but it also entails major implications in respect of what will be required of institutions before this potential can be fully realised. It should, for example, be possible one day for staff and users to have access to a 'clump' of cross-searchable collection strength indices based at COPAC, the National Libraries, and members of CAIRNS, M25 and RIDING. These indices could be built automatically at marginal cost by library management systems operating in real time as institutional cataloguers added records, could offer up to the minute data to both users seeking materials and staff engaged in deep resource sharing, and could theoretically allow filtering for uniqueness (regardless of numerical strength), language, research or teaching code, location, access category and other relevant elements. As with other automated approaches, however, full functionality will only be possible if institutions are willing to undertake significant long-term work on metadata, if common classification and subject schemes are adopted, if institutions will add additional collection strength data to item records, if they will pay for the creation and maintenance of associated indices, and so on. Exactly how the distributed index described above would be designed, and what the specific implications for institutions would be requires more detailed work as described below under 'Recommendations'. There may be value in considering examining this approach in the arena of e-resources in the first instance, there being less of a problem with legacy metadata in this area.

It is possible, of course, that there are various intermediate solutions that will entail less in respect of requirements on institutions, and that one or more of these may be found attractive because of this, despite a likely reduction in the value and accuracy of the data. Improving iCAS algorithms for mapping materials classified against a

non-iCAS scheme or not classified at all to an iCAS scheme may be one such approach. Equally, it may be that such a reduction is acceptable in some cases where data accuracy is sufficient to meet requirements and the method employed is cheaper than a more accurate assessment would be. Determining the number and precise nature of such intermediate solutions is beyond the scope of this evaluation, however. Again, these would be drawn out by the way forward proposed below under 'Recommendations'.

## **Recommendations**

**4.1** Aim to build a consensus on a UK-wide collections mapping policy, this to be based initially on a mix of automated and non-automated methods but to have the long-term goal of optimising the use of automated methods where these are the most cost-effective and affordable option

**4.2** Inform this process by funding research into the issues listed below, either through a single project or through a co-ordinated group of projects:

- Research on actual requirements of funding bodies and policy makers, inter-institutional groups (like CURL, SCURL, M25, RIDING, CALIM), institutions, collection managers, and users in respect of access to materials and resource sharing, and on the best way of building these into either automated or non-automated assessment methods, bearing in mind the various elements drawn out in the table in Appendix A.
- An examination of alternative manual and automated approaches to meeting these requirements, their limitations, and the best means of overcoming these. This to include an examination of costs against benefits in relation to each approach, involve OCLC, system vendors and open source experts, and include a consideration of a UK-wide deal with OCLC iCAS.
- An examination of associated requirements on institutional policies and practices and of the practicalities and value of implementing these as widely as is necessary in UK institutions, taking into account considerations such as the metadata creation and maintenance requirements of a particular approach.

One possible approach would be a research project that seeks to combine the 'constrained professional judgement approach' recommended by SCONE in Scotland as an approach to existing collections and an automated approach based on distributed indices for newly acquired materials. Comparing this with an iCAS based approach would permit research into the various elements outlined above but only require 'full disclosure', adopting a single scheme, adding additional metadata elements, and so on, in respect of newly acquired material. Possible participants would include CURL, SCURL, M25, RIDING and CALIM institutions, the DNER and the RDN. DNER and RDN might have an interest because of the possible usefulness of distributed collection strength indices as a means of allowing users to navigate the JISC Information Environment.

**4.3** Determine at the end the best way forward in respect of a UK-wide policy, bearing in mind all of the following:

- The requirement of all of the key stakeholders, including funders and users
- The implications of a particular approach in terms of institutional and inter-institutional practices (e.g. the need to do things such as implement Full

Disclosure, adopt a common subject scheme, agree and apply common standards generally, build new indices, do deep resource sharing, agree relevant inter-access policies, and so on)

- The likelihood that institutions will implement and maintain these practices over time, given staffing and funding implications and other relevant consideration.

**Dennis Nicholson**  
**26.04.02**

<b>Appendix A</b>				
<b>Optimum Position</b>	<b>Current Position</b>	<b>Implications</b>	<b>Solution Paths</b>	<b>Additional Comments</b>
All items in all collections catalogued and online; none have cataloguing backlog when collection strength measurement snap shot is taken	This is not true of most collections in the UK and it is often the older, most important collections that are not catalogued, a fact reflected in the figures from the CURL study	Manual work needed to complete the assessment; this likely to be of variable level across the institutions; resultant reduction in objectivity and increase in costs over automated costs	A gradual move towards complete inter-compatible catalogues of all materials at all UK sites; Value in considering e-materials separately to obtain earlier improvements in metadata in e-materials	Automated analysis likely of most value currently to libraries like Hull with 100% online catalogues, or those with only easily assessed older collections offline; Whether costs within budget a factor
Individual items are distinguished not just by subject, but by other criteria are <i>known to be</i> needed by users and staff (e.g. whether teaching or research, what language, intended use, current or historical subject relevance etc.)	There is little or no machine processable data of this kind in records of items in either the CURL iCAS study or elsewhere, and more information is required on actual user and collection manager needs, particularly in respect of collaborative collection development initiatives	Automated methods can indicate that collections are numerically 'strong' but users and staff likely to need to know specifics - e.g. a user seeing a 'strong' research collection may be misled by a 'strong' coding based on a general count	Research on actual user and collection manager requirements and on the best way of building these into automated assessment (e.g. by libraries coding particular class mark ranges as mainly bought for research or for teaching or both)	If the main function of a collective is to build and promote the use of members collections mainly for either research or teaching, but not both, this element is less of a problem (improvements still required, however).
Analysis software can breakdown based on these additional characteristics; same descriptive terms used for these by all and chosen to convey limits of usefulness objectively	Online catalogue data does not usually entail the terms, iCAS software cannot use them, there is no gauge on limits of usefulness, and no data on how best to ensure objective description	Manual assessments may in some cases provide better guidance currently to users and staff if methods of ensuring objectivity of these are first agreed and then applied consistently	Research as described in the column above, but taking the points made here into account; software developments to allow new data to be reliably extracted.	Numerical assessment is objective in the sense that it counts exactly what it counts. However, this objectivity can be lost if data is then interpreted subjectively or described in a way that implies this.

Optimum Position	Current Position	Implications	Solution Paths	Additional Comments
Collection analysis software allows unique material to be highlighted even in small collections, preferably based on a unique identifier such as ISBN so that users and staff can tell if numerically weak collections have unique material	iCAS provides uniqueness data, but not to title level and likely to be affected by accuracy of comparison algorithms and of problems noted elsewhere in this chart (such as differences in metadata standards).	Uniqueness data should be recorded and made available to users and staff; some manual evaluation of automated uniqueness assessment is probably advisable	Inter-compatible metadata using unique identifiers in all cases and software functionality able to show uniqueness accurately at title level based on unique identifiers	In an ideal world this would be 'up to the minute' information, available to users and staff soon after a new unique item was added to a collection
Collection analysis software can indicate numerically small but significant strengths	iCAS shows uniqueness but not significance; metadata does not record significance; significance requires human assessment	Some non-automated assessment will always be required for this, but uniqueness measures at least alert users and staff to check for significant material	Needs agreed definition of significance; agreed method of describing and recording it; software that can process the metadata	Given that manual assessment a requirement here, a simple staff coding that indicates significance may be sufficient and can be assigned without automation
All libraries use one version of an iCAS compatible classification scheme for collections and items, apply it accurately 'as is' to all items at appropriate agreed granularity level, apply the same number to the same item, form and record the number correctly, update the scheme frequently	As HILT <sup>3</sup> has shown, a wide variety of schemes are in use across the UK, there is currently no standardisation of approach even where the same scheme is used, some use no scheme, or a scheme that is not iCAS compatible, some do not apply the scheme to all items	Automated methods will not be applicable to all libraries and will entail a (currently incalculable) margin of error that will require manual intervention at a local level and that will make inter-institutional comparisons suspect	Adopt a common subject scheme and apply it in a common way to all items everywhere <sup>4</sup> ; have common training scheme; perhaps investigate whether it is possible to usefully gauge margin of error in current situation based on things like % of stock covered; edition differences etc.	This is a major barrier to the use of automated methods in or across libraries who fall short of the situation described under 'optimum position' in this row and, hence, a significant difficulty for the use of automated methods in the UK as a whole

<sup>3</sup> See <http://hilt.cdli.strath.ac.uk/>

<sup>4</sup> Or, at worst, a scheme entailed in any inter-scheme mapping solution agreed in HILT Phase II

<b>Optimum Position</b>	<b>Current Position</b>	<b>Implications</b>	<b>Solution Paths</b>	<b>Additional Comments</b>
Data gathered beneficial and relevant to both users and collection managers	Data on actual requirements of users and staff in respect of collection analysis and mapping is sparse at best.	Unless we know in detail what we want to do with collection analysis and mapping data, we cannot readily assess what data we need, how it may best be gathered, and how much effort and expense it is worth committing to it.	Conduct research to determine - in respect of all levels of user and all types of institution - what the requirement is, taking into account a range of factors, but including the view if users, staff, institutions, inter-institutional bodies like CURL and SCURL, and funders and national policy makes.	A major issue of high priority - should be tackled at an early stage
Total cost is within the means of all participants	Not known in detail, but cost of automated mapping through methods such as iCAS would probably be regarded as too high by many.	More information is required, but the long-term possibility that data of this kind might be automatically available from local library systems should be investigated	Research the requirements in respect of having this kind of data available from library systems; talk to system vendors about the best way to implement this.	There is no reason why, in the long term, up to the minute information of this kind should not be available a marginal cost in both central systems like COPAC and distributed systems like CAIRNS, RIDING, M25
Benefits outweigh costs, even if cost within budget	Not known, but clearly a consideration	Time, effort, funds spent in this area should focus on research to minimise the potential for waste caused by lack of hard information.	Requirements analysis should consider costs against benefits as a factor	Could make the 'minimalist' SCONE approach best in the short term

<b>Optimum Position</b>	<b>Current Position</b>	<b>Implications</b>	<b>Solution Paths</b>	<b>Additional Comments</b>
It is measured against a baseline such as Worldcat and the baseline itself covers all publications in all countries, all cultures, all languages	This was not done in the present study, but if the aim is to alert users and staff to the strengths of UK collections in a global context, and of the need to look elsewhere for materials in some instances, then it is a key element for the future; Even Worldcat would not claim the ideal level of comprehensiveness	Users and staff may know collection weaknesses and strengths in relation to the collections of participants but will be unaware of whether material elsewhere is worth pursuing.	Aim for a global approach as far as possible (e.g. all major research libraries world-wide); include a comparison with major databases like Worldcat if possible	Requirements and cost-benefit analyses relevant here.
On order items included	They are not included	Early classification estimate	Include on order items	Importance depends on depth of resource sharing
Only items in collections accessible to the user who is searching are shown in strength measurements for individual users	Not directly available through automated methods but could be	Useful for users to see strengths even if they can't have access but also useful for users to know they can't have access	Include accessibility data; determine best way of approaching this	Not of major importance but useful to the user
There is a flexible approach to age analysis	Level of flexibility unknown.	A single approach may not be useful to all	Age data should be recorded in raw form to allow maximum flexibility	Level of need for flexible approach requires examination.
Apply different class numbers for items covering more than one subject area	Not known but probably not widely applied	Unknown, may or may not have significant bearing on strength measurements	Determine how significant this is; apply in future if it is significant considering both costs and benefits	Possibly not a major issue but should be examined to confirm

<b>Optimum Position</b>	<b>Current Position</b>	<b>Implications</b>	<b>Solution Paths</b>	<b>Additional Comments</b>
The objectivity of the data is not undermined by the claims made	Numbers are objective in the sense that they measure what they measure. Noting that three collections have 100, 200, and 300 items is objective, indicating that the third of these is a 'strong' collection may not be - e.g. strong in relation to what, for whom, for what purpose are considerations.	Objective measurement methods can be undermined by subjective approaches to description.	Determine what can be measured objectively, how it can best be objectively described, and what the limits of its use are for both users and staff	An important aspect of a requirements analysis
Title matches are of a very high accuracy	Level of accuracy in iCAS unknown	Overlap and uniqueness measures flawed if not	In the long term, a unique identifier for every item	Manual methods cannot normally provide this level of detail at all so automated methods best here already
Up to the minute data on strength, overlap, and uniqueness allowing access down to individual and identifiable titles across all participating institutions	Even iCAS can only offer occasional (and relatively expensive) snapshots at a given point in time and does not provide title level information on overlap and uniqueness	iCAS data insufficient in itself to support implications of 'deep resource sharing' for either staff practising it or users living with it. The long term aim should be something better	Aim long-term for cross-searchable 'collection strength' indices at each participating institution, these to include subject, unique identifier, language, research or teaching code, location, access category	With Z39.50, this could be done now given the existence of appropriate and inter-compatible metadata at participating sites - as ever, the problem is in (or not in) the metadata.

### **5.12 Acronyms**

ACAS	Automated Collection Analysis services
CURL	Consortium of University Research Libraries
iCAS	Interactive Collection Analysis System
ISBN	International Standard Book Number
ISSN	International Standard Serial Number
LC	Library of Congress
LCCN	Library of Congress Control Number
NHM	Natural History Museum
NLM	National Library of Medicine
OCLC	Online Computer Library Center
RSLG	Research Support Libraries Group
RSLP	Research Support Libraries Programme
SOAS	School of Oriental and African Studies

### **5.13 List of Key Contacts**

CURL	Anne Mealia, former CURL Database Officer (until September 02)  Chris Bailey, former CURL Executive Secretary, currently Director of Library Services at the University of Glasgow  Author of final report – Marie-Pierre Détraz, CURL Executive Secretary since February 02
RSLP	Ronald Milne, Programme Director
OCLC	Glenda Lins, Product Manager, OCLC Lacey Product Center  Sally Loken, OCLC ACAS Consultant
Participating Libraries	Imperial College (CURL member) – Clare Jenkins, Director of Library Services  Natural History Museum – Neil Thomson, Head of Systems and Central Services

School of Oriental and African Studies – Keith Webster, Librarian and Director of Information Services

University of Edinburgh (CURL member) – Richard Ovenden, Special Collections Librarian

University of Hull – B. A. Towler, Head of Digital Library Development, Academic Services Libraries

University of Liverpool (CURL member) – Frances Thompson, University Librarian

External Evaluator

Dennis Nicholson, Director of the Centre for Digital Library Research, University of Strathclyde

END OF APPENDICES