

JISC DEVELOPMENT PROGRAMMES

Project Document Cover Sheet

Final Report

Project

Project Acronym	CC-interop	Project ID	
Project Title	COPAC/Clumps Continuing Technical Cooperation Project		
Start Date	1 May 2002	End Date	30 April 2004
Lead Institution	London School of Economics		
Project Director	Jean Sykes, Librarian & Director of Information Services		
Project Manager & contact details	<p>John Gilby (WP A & C) M25 Systems Team Project Manager, LSE Library, 10 Portugal Street, London WC2A 2HD tel: 020-7955-6451 email: j.gilby@lse.ac.uk</p> <p>Sean Dunne (WP A) MIMAS, Manchester Computing, Kilburn Building, The University of Manchester, Oxford Road, Manchester M13 9PL tel: 0161 275 6065 email: sean.dunne@man.ac.uk</p> <p>Gordon Dunsire (WP B) Depute Director Centre for Digital Library Research, Livingstone Tower, Strathclyde University, 26 Richmond Street, Glasgow G1 1XH tel: 0141 548 4680 email: g.dunsire@strath.ac.uk</p>		
Partner Institutions	MIMAS, University of Manchester Library, London School of Economics and Political Science CDLR, Strathclyde University		
Project Web URL	http://ccinterop.cdli.strath.ac.uk/		
Programme Name (and number)			
Programme Manager	Chris Awre		

Document

Document Title	Final Report		
Author(s) & project role	John Gilby, Project Manager (WP A & C) Gordon Dunsire, Project Manager (WP B)		
Date	June 2004	Filename	CCiFinalReportVersion1.doc
URL	http://ccinterop.cdli.strath.ac.uk/documents.htm		
Access	✓ Project and JISC internal	✓ General dissemination	

Document History

Version	Date	Comments
0a	May 2004	First Draft
0b	June 2004	Second Draft
0c	June 2004	Final Draft
1	June 2004	First Issue



**COPAC/Clumps Continuing
Technical Cooperation Project**

Final Report

Version 1

June 2004

Authors

John Gilby (London School of Economics),
Gordon Dunsire (University of Strathclyde)

Lead Institution

London School of Economics and Political Science

Contact Details

Director: Jean Sykes, Librarian and Director of Information
Services
Address: LSE Library, 10 Portugal Street, London WC2A 2HD
Tel.: 020 7955 7218
Fax: 020 7955 7454
Email: j.sykes@lse.ac.uk

Table of Contents

Acknowledgements	iii
Executive Summary	1
1. Background	3
2. Aims and Objectives	4
2.1. Work Package A	4
2.2. Work Package B	6
2.3. Work Package C	8
3. Methodology	9
3.1. Work Package A	9
3.2. Work Package B	11
3.3. Work Package C	16
4. Implementation	18
5. Outputs and Results	19
5.1. WP A	20
5.2. WP B	22
5.3. WP C	29
6. Outcomes	31
7. Conclusions	32
8. Implications and Recommendations	33
9. References	35
Appendix A	37
Appendix B	41

ACKNOWLEDGEMENTS

The CC-interop Project received funding from JISC following a free-standing bid submitted to the JISC Committee for the Information Environment.

Partner staff working directly on the project were John Gilby and Fraser Nicolaides (M25 Systems Team), Dr. Shirley Cousins and Ashley Sanders (MIMAS), Gordon Dunsire and George Macgregor (CDLR) and Bridget Towler (RIDING). In addition, Helen Booth and Dick Hartley (CERLIM/MMU) were responsible for the WP C User Behaviour Study.

The Project wishes to acknowledge the contributions made by the partner institutions with respect to staff accommodation, secretarial, administration and IT support. Also to members of the project Steering Group for their interest and guidance throughout the duration of the project.

Particular thanks go to non-CC-interop clumps in expressing their views on technical issues for a number of the Work Package A tasks. Specifically, Rob Bull of Crossnet Systems Limited, has been very generous in sharing his extensive knowledge on Z39.50 issues.

Thanks are also extended to those who attended events in connection with the cataloguing and indexing guideline tasks of Work Package B, and also to those volunteers that gave up their time to attend user sessions or focus groups in connection with the user behaviour study under Work Package C.

Finally, the authors wish to acknowledge the assistance given in the preparation of this report by George Macgregor (CDLR).

This page intentionally blank.

EXECUTIVE SUMMARY

As far as is known, CC-interop was the first project of its kind anywhere in the world and still is. Its basic aim was to test the feasibility of cross-searching between physical and virtual union catalogues, using COPAC and the three functioning "clumps" or virtual union catalogues (CAIRNS, InforM25, and RIDING), all funded or part-funded by JISC in recent years. The key issues investigated were technical interoperability of catalogues, use of collection level descriptions to search union catalogues dynamically, quality of standards in cataloguing and indexing practices, and usability of union catalogues for real users. The conclusions of the project were expected to, and indeed do, contribute to the development of the JISC Information Environment and to the ongoing debate as to the feasibility and desirability of creating a national UK catalogue. They also inhabit the territory of collection level descriptions (CLDs) and the wider services of JISC's Information Environment Services Registry (IESR). The results of this project will also have applicability for the common information environment, particularly through the landscaping work done via SCONE/CAIRNS. This work is relevant not just to HE and not just to digital materials, but encompasses other sectors and domains and caters for print resources as well.

Key findings are thematically grouped as follows:

System performance when inter-linking COPAC and the Z39.50 clumps

- The various individual Z39.50 configurations permit technical interoperability relatively easily but only limited semantic interoperability is possible
- Disparate cataloguing and indexing practices are an impairment to semantic interoperability, not just for catalogues but also for CLDs and descriptions of services (like those constituting JISC's IESR)

Creating dynamic landscaping through CLDs

- Routines can be written to allow collection description databases to be output in formats that other UK users of CLDs, including developers of the JISC information environment, can harvest and use

Searching a distributed (virtual) catalogue or clump via Z39.50

- Use of Z39.50 to Z39.50 middleware permits a distributed catalogue to be searched via Z39.50 from such disparate user services as another virtual union catalogue or clump, a physical union catalogue like COPAC, an individual Z client and other IE services
- The breakthrough in this Z39.50 to Z39.50 conundrum came with the discovery that the JISC-funded JAFER software (a result of the 5/99 programme) meets many of the requirements and can be used by the current clumps services
- It is technically possible for the user to select all or a sub-set of available end destination Z39.50 servers (we call this "landscaping") within this middleware

Comparing results processing between COPAC and clumps

- Most distributed services (clumps) do not bring back complete results sets from associated Z servers (in order to save time for users)
- COPAC on-the-fly routines could feasibly be applied to the clumps services
- An automated search set up to repeat its query of 17 catalogues in a clump (InforM25) hourly over nearly 3 months returned surprisingly good results; for example, over 90% of responses were received in less than one second, and no servers showed slower response times in periods of traditionally heavy OPAC use (mid-morning to early evening)

User behaviour when cross-searching catalogues

- The importance to users of a number of on-screen features, including the ability to refine a search and clear indication that a search is processing
- The importance to users of information about the availability of an item as well as the holdings data
- The impact of search tools such as Google and Amazon on user behaviour and the expectations of more information than is normally available from a library catalogue
- The distrust of some librarians interviewed of the data sources in virtual union catalogues, thinking that there was not true interoperability

1. BACKGROUND

During 1998, four projects began as part of the *Large Scale Resource Discovery* strand of the JISC eLib Phase 3 Programme. These have become colloquially known as “the clumps”, and comprise CAIRNS, InforM25, Music Libraries Online (MLO) and RIDING¹. All four projects still exist, either as full functioning services or as advanced, well-used prototypes. In all cases, the original eLib grant has finished but development effort and maintenance continues at varying levels, either funded by local consortium subscriptions (CAIRNS, M25 and RIDING) or by small amounts of funding from elsewhere.

COPAC <http://copac.ac.uk> is established as a core JISC service at MIMAS, and is a significant cornerstone of the JISC Information Environment. Based on the consolidated catalogue database of the CURL members, COPAC offers a single interface to the bibliographic records of many major university research libraries in the UK and Ireland, plus the British Library and National Library of Scotland.

The CC-interop project builds on the results of JISC’s eLib Phase 3 programme in the area of resource discovery and complements the JISC Information Environment, particularly the Infrastructure Programme, which includes investigation of Z39.50. CC-interop also enhances the “distributed” thread of the JISC Information Environment, in that it aims to bring together, in a virtual *modus operandi*, distributed catalogues to facilitate richer search and retrieval possibilities for users. This, in turn, exemplifies the JISC vision, stating that, “it is not a centralised service and does not rely on a single dedicated entry point”.

CC-interop also resonates with the outcomes and recommendations of the RSLG (Research Support Libraries Group) Final Report (2003), and continues to resonate with the aims of the proposed Research Libraries Network (RLN). RLN recognises that libraries might in future collaborate through shared access and collection management in order to support the UK research community as a whole. Linking high-quality online catalogues together can be seen as an essential precursor to the further development of shared access and borrowing agreements between HE libraries, to the development of shared collection development and management policies, and to more document delivery services in support of researchers.

¹ At the beginning of this project, RIDING was a full partner, assisting with the tasks in Work Package B. During the summer of 2003, RIDING completed their work on Work Package B and, with the agreement of the remaining partners, opted out of the Project due to organisational changes at the University of Hull. CAIRNS is maintained by CDLR based in the University of Strathclyde and InforM25 is maintained by the M25 Systems Team based in the LSE.

2. AIMS AND OBJECTIVES

The primary aims of this project were split into three groups, two (Work Packages A & B) due to the organisation of the teams carrying out the work, and thirdly, a user behavioural study (WP C).

In WP A the M25 Systems Team at the LSE and MIMAS at the University of Manchester investigated the feasibility of inter-linking between a very large physical union catalogue and a large virtual union catalogue. A number of issues were included in this investigation such as comparative speed of searching, de-duplication, results ranking and also comparing the accuracy both of the records themselves and the results.

The WP B work undertaken by Centre for Digital Library Research (CDLR) and RIDING staff based in the University of Hull looked at collection level description schemas in relation to both the clumps and COPAC. Issues such as target selection in clumps and developing guidelines for cataloguing and indexing practices were also included.

WP C analyses the behaviour of users when they search COPAC and the "clumps", WP C contributing to our common understanding of how JISC's IE functional model might be instantiated in an operational environment. Covering such issues as wide-area item-level discovery and collection descriptions, the component parts of CC-interop fit very closely with the IE functional model, and WP C results have produced issues identified by users, or observed in their behaviour, which can inform the development of the IE as well as the National Catalogue which will be part of the IE.

2.1. WORK PACKAGE A

This work package aimed to build on the existing experience of COPAC and the M25 Systems Team by exploring how they could technically inter-operate. The experimental COPAC v.3 <http://copac.ac.uk/msgw/wzgw> has shown that it is possible to cross-search a large union catalogue such as COPAC and remote catalogues such as those of the National Libraries of Scotland and Wales via Z39.50. Work by the eLib Clumps has shown that Z39.50 can be used to cross-search several distributed library catalogues.

The feasibility of cross-searching COPAC and the M25 libraries was investigated from two different starting positions. Firstly a version of InforM25 was built which included COPAC as one of its Z39.50 targets, and secondly, a Z39.50 target for InforM25 was developed and added as a target to be searched by COPAC v.3. In both cases we explored several areas which are relevant to cross-searching the different resources.

In addition, there are a number of issues/areas that are relevant to both the options above, and to COPAC and the clumps as stand-alone services, that were jointly investigated and solutions shared. Many of these will also provide input to a potential UK National Catalogue.

The tasks for this work package are detailed in the following list:

1. Develop a demonstrator version of InforM25 with COPAC as an additional Z39.50 target, carrying out thorough testing to ensure the Z-searching is working with optimal Bib-1 attribute values and other settings.
2. Investigate the feasibility (& practicality) of building a Z39.50 target for InforM25 – this task will investigate possible technical models to transform a clump into a Z-target, primarily with InforM25, but also looking in detail at the other clumps.
3. If feasible, carry out work with the InforM25 system to create a Z-target.
4. Add the InforM25 Z-target to COPAC v3 and carry out testing – the exact nature of the testing will depend on the technical solution employed. Also, consideration will be given as to the *value* that a clump offers COPAC (in an UKNUC context) as opposed to COPAC Z-searching the clump constituent library catalogues directly.
5. Look into common areas such as de-duplication, results set sorting and ranking – An investigation into the different methods employed by COPAC and the clumps and into the feasibility of developing common routines/processes in the different architectures giving consistency to the user within the current differing user interfaces. A key issue to consider will be where or how to decide on an *authoritative* record when records are combined from the disparate sources. At what point results are processed will also be investigated, e.g. by COPAC or by the clump prior to sending to COPAC for the Task 4 model.
6. Carry out a critical review of results obtained for the same searches, firstly using InforM25, searching the M25 CURL libraries, and secondly, via COPAC. The results will be compared at the user interface level, i.e. comparing what the user would see on the two services. If deemed necessary, direct OPAC searches can also be used as a reference. This review will consider a number of issues, some raised in the UKNUC Feasibility Study Final Report to look at the quality of the returned records (if the same MARC fields & data are shown in the results by looking at the same item records presented in COPAC and InforM25), the consistency of results sets and, where possible, recall and precision of the results sets.
7. Investigate response times for Z39.50 searching – The UKNUC Feasibility Study Final Report comments on search performance via Z39.50 and all Partners have either experienced variation in search times or been faced with general comments that Z-searching is slow. Technical investigations will be carried out to identify factors that affect the response times and, depending on the findings, possible remedial activities undertaken. There will also be an investigation of work undertaken by members of the ZIG on this issue. An element of the Task may also look at performance issues of local OPACs when the library system is subjected to multiple Z39.50 searches.

Partners working on WP A were successful in carrying out all of the above tasks though for convenience, some were grouped together.

Significantly, it was discovered that the JAFER² project software had the technical capability to act as Z39.50 to Z39.50 middleware and a copy of the software was obtained for use on CC-interop. Use of the JAFER software was entirely successful in demonstrating that building a distributed catalogue that was itself a Z-target was technically and practically feasible.

Task 1 was combined with elements of Task 6 to compare searches via InforM25 and COPAC and to carry out detailed tests to highlight differences in search attribute and indexing configurations.

Tasks 5, 6 & 7 were also worked on together for the convenience of the staff working on this work package.

2.2. WORK PACKAGE B

Work Package B (WPB) was predominantly undertaken by the Centre for Digital Library Research and in broad terms investigated the use of collection-level description schemas in relation to both the clumps and COPAC, including issues such as target selection in clumps and developing guidelines for cataloguing and indexing practices. This entailed the development of SCONE as a 'dynamic clumping' facility for cross-searching institutional catalogues in CAIRNS and COPAC, and sought to address the issue of the same metadata being available in two or more catalogues arising from the overlap between COPAC and the clumps. The augmentation of clump coverage was an integral component of this and therefore dictated that WPB extend the collections coverage in CAIRNS and RIDING, and encompass COPAC both as a service and a set of collections. The export of collection descriptions from SCONE into different formats was also investigated by WPB and appropriate facilities have been instantiated within the SCONE technical architecture.

An additional aim of WPB was to enhance intra- and inter-clump interoperability by compiling guidelines for best practice in respect of cataloguing and indexing standards in clumps and outlining a medium-term strategy for implementing them.

It should be noted that it was the intention that the RIDING clump work closely with the CDLR, with respect to WPB. However, whilst RIDING was involved initially, their participation in the project ceased due to unforeseen circumstances, 6 months into the project schedule. Though this clearly had implications for the aims and objectives of WPB - and given the junior nature of RIDING's relationship to the project - CDLR was able to assume control of RIDING's work and in many cases produced enhanced manifestations of the original deliverables, particularly with respect to deliverables B3 and B2, as per the project plan.

² The JAFER Toolkit Project was funded by JISC under the 5/99 Programme. See <http://www.lib.ox.ac.uk/jafer/index.html>

The distinct aims of WPB were as follows:

- Investigate and specify collection description standards requirements for clumps as JISC Information Environment nodes.
- Enhance the coverage of clumps services by extending collections coverage in two clumps and encompassing COPAC both as a service and a set of collections.
- Enhance intra- and inter-clumps interoperability by compiling guidelines for best practice in respect of cataloguing and indexing standards in clumps and outlining a medium-term strategy for implementing them.
- Investigate the landscaping of mini-clumps involving services and collections from two clumps and collections from COPAC.
- Enhance the structure of the SCONE database to take account of the full requirement if necessary.
- Enhance the test-bed collections adding and amendment facility in the associated SCAMP staff portal to take account of the new structure.
- Agree a common subject scheme or schemes for describing collections based on HILT (1) outcomes (probably DDC or LCSH or both) and - if appropriate - HILT (2) outcomes.
- Create a new regional database and updating portal appropriate to the specific requirements of RIDING, populate it with a set of collection descriptions appropriate to the RIDING clump, and link it within the RIDING service.
- Provide guidelines for the use of the staff updates portal within RIDING.
- Create a set of records that can be added to the databases of both clumps (and others), describing COPAC collections to a sufficient degree to allow COPAC to be appropriately landscaped, possibly based on the outputs of the current CURL/RSLP iCAS Collection Analysis Project into this and on the collection strengths data from CURL libraries already in CAIRNS.
- Add them to both databases, enhancing coverage further.
- In CAIRNS, encompass COPAC appropriately as a service within the environment, so that it will be added to dynamically generated landscapes in appropriate circumstances.
- Investigate simulating the same process for RIDING, but using the CAIRNS Z39.50 client.
- Investigate the collection descriptions-generated dynamic landscaping of 'mini-clumps' that offer cross-searches of a mix of services from RIDING and CAIRNS, together with COPAC where appropriate.
- Write routines that will allow the collection description databases to be output in formats that other UK users of collection descriptions, including the JISC Information Environment, can harvest and use.
- Conduct extensive tests of the databases and their operation to determine whether or not the structure and content of the databases are adequate to

the requirement. Enhance the structure and content if this proves not to be the case.

- Provide guidelines to other project participants for implementing the agreed requirement in respect of collection descriptions within their own environments, either by utilising the CAIRNS/RIDING approach or by utilising some other agreed approach.
- Write a report on the results that can be utilised by others, including JISC services.

WPB achieved all the aforementioned aims, some to a greater or smaller degree than originally envisaged.

Difficulties with the contribution of RIDING have affected some aims. With respect to the creation of a regional database, no link was made to the RIDING (public) service. Little progress was made with encompassing RIDING as service within the CAIRNS environment, as achieved with COPAC. With reference to the enhancement of intra- and inter-clumps interoperability via the creation of guidelines for best practice, the outcome has been a set of recommendations for improving interoperability in a much wider context, rather than a specific set of guidelines.

No progress was made with using the outcomes of the CURL/RSLP iCAS project.

2.3. WORK PACKAGE C

As noted above, CERLIM at Manchester Metropolitan University were sub-contacted to undertake the user behaviour study for the project. CERLIM's role was to investigate the behaviour of typical end-users in their searching of union catalogues, regardless of whether the catalogues were physical or virtual. There was emphasis placed on studying how the user carried out searches of the clumps and COPAC services, and the reasons behind their actions. Looking for common behaviour and/or difficulties between the different services was also part of the requirement. Investigation of search effectiveness was not within the remit of WP C as this was covered within WP A.

This work package was completed and further details can be found in later sections of this report.

3. METHODOLOGY

From the outset, the project centred on specific issues concerning union catalogues of different technical architectures and how these might interoperate successfully. The work packages were drawn up at the bid stage to use the relevant skills and experience of the different partners advantageously and to carry further the findings arising from the eLib clumps projects.

Thus, the M25 Systems Team and MIMAS with their knowledge of Z39.50, interoperability and union catalogue design, took responsibility for tasks under WP A. CDLR and RIDING during their respective eLib periods, gained considerable experience with collection descriptions and particularly on using these to define the information landscape and so took on the WP B tasks. All the partners agreed that a more objective approach for the user behaviour study was necessary and that it would be inappropriate for the work to be undertaken by one or more of the partners. An external consultancy was preferred and CERLIM selected to carry out the work.

Brief details of the differing methodologies used within the work packages follow.

3.1. WORK PACKAGE A

To investigate interoperability between union catalogues of distributed and non-distributed architectures, WP A concentrated firstly on whether the two could actually be connected (i.e. adding a clump to COPAC and vice versa); and, secondly, investigated relevant issues such as searching performance, landscaping and results issues.

Programming effort was available at both MIMAS and the LSE enabling COPAC to be included as a target on InforM25, and for a new version of InforM25 capable of acting as a Z-target, to be built at MIMAS and searched by COPAC.

In addition to using the knowledge from within the project team, throughout the project period several other clumps services and experts from the UK and beyond were consulted as well as two vendors who are active in the area of distributed searching. In particular, the views of Carrol Lunau (National Library of Canada), Rob Bull (Crossnet Systems Ltd.) and Matthew Dovey (University of Oxford) were sought as these colleagues have extensive experience of Z39.50 and are actively involved in relevant groups and standards such as the Bath Profile and the Z39.50 Implementers Group.

InforM25 Copy

A copy of InforM25's "Search catalogues" service was created (referred to as CC25) for the purposes of the project and COPAC added as a Z-target. Other M25 targets on CC25 used the same configurations as the live service for comparative purposes. Additionally, an independent Z-client was utilised to perform searches thus allowing specific semantic definitions to be tried that were not available on CC25.

JAFER

A copy of the JAFER software was obtained from the JAFER website. Being unfamiliar with some of the technologies used by JAFER the project acknowledges the help freely given by the JAFER team. During exploration of the software a number of bugs were found and reported. Again, the project appreciated the prompt response of the JAFER team in resolving these problems as they emerged.

A number of developments were carried out by MIMAS to facilitate the use of JAFER for the purposes of the project.

The logging facilities were extended significantly as we found the logging built into JAFER was not detailed enough for our purposes. It did not allow us to see clearly what communication was taking place between JAFER and the individual library Z39.50 targets. In particular we extended the logging of the Search and Present request and response APDUs. This meant writing some new Java code as well as making modifications to the JAFER code itself. The new Java code has been passed back to JAFER.

The M25 Systems Team provided a set of Microsoft Excel tables detailing how the InforM25 system was set-up. These tables were converted to plain ASCII text and then a perl script was written to convert them to the XML format required by JAFER. These tables provide the best practice settings for connecting to the individual InforM25 library catalogues.

JAFER handles the Z39.50 query internally as XML. It allows the query to be transformed using XSLT before it is forwarded to each target. A range of XSLT programs, which may be looked on as stylesheets, were developed for transforming search queries as required by each library's Z39.50 target. The specification for each stylesheet was taken from the InforM25 set-up files. The stylesheets are integrated into the JAFER target to ensure each query is presented to a particular target in the form most suited to it. For example, most of the InforM25 targets accept a journal title search on Use attribute 33, whereas COPAC uses Use attribute 4 for all types of title search and requires an additional search term of "periodical" against Use attribute 1031 (Material type) to limit the search to journals. Therefore XSLT stylesheets were programmed to remove the search against Use attribute 1031 and change Use attribute 4 to 33 when a journal title search was required. Similarly, other attributes were removed and added as required by the InforM25 specifications.

One option for the Z-target is to simply configure it to search all InforM25 library catalogues. However this may not always be appropriate, for example a user may wish to restrict a search to their local area ensuring easier access to materials identified by searching the Z-target. To explore this a number of different pre-defined subsets of libraries (known as concatenations) were established enabling all InforM25 contributors to be included in a search, or to restrict the search to particular sub-sets. The following concatenations were established using InforM25 defined library groups, the last three corresponding with the InforM25 zones:

All InforM25 Libraries, North London, South London, Inner London

For testing purposes we also established two concatenations that related the libraries to COPAC. If the InforM25 Z39.30 target was to be used to search libraries alongside the COPAC database then it was necessary to exclude those libraries already represented on COPAC (as of July 2003), to avoid duplication of results from these institutions. The two additional concatenations were simply:

M25 Libraries on COPAC

M25 Libraries not on COPAC

COPAC Interface Copy

To enable the search testing of the JAFER InforM25 Z39.50 target a new version of the COPAC V3 web interface was created that was isolated from the live COPAC V3 service. This enables independent logging of the JAFER system and allowed us to amend the interface without impinging on COPAC users. Some development work was required to enable us to test it with JAFER. Changes to the COPAC Z39.50 origin and web interface were required to enable the COPAC record to display the individual source library for each record retrieved. The source library is obtained from the database name that JAFER includes in each record. This provided a front-end for JAFER enabling the searcher to use the COPAC interface to search the InforM25 libraries.

Results and Display Issues

These WP A tasks looked at search result manipulation and covered such areas as: combining distributed search results; de-duplication/matching of combined result sets; consolidation of records (creating a single bibliographic record); ranking; sorting, record suppression (e.g. how to cope with records not suitable for user display; and other related result processing issues.

The current COPAC service was looked at and the various methods of search result manipulation and associated display issues detailed. Similar issues were then reviewed with respect to the distributed environment and an initial review of whether the COPAC techniques would be technically possible within a multi-stage distributed environment considered.

Views were also obtained on the same results processing issues from other clump service providers and a small number of vendors with portal-type products.

3.2. WORK PACKAGE B

Collection description standards requirements

Deliverable B1 built on 'An Analytical model of collections and their catalogues', a study of collections and collection description conducted by Michael Heaney on behalf of the UK Office of Library and Information Networking with support from OCLC (Heaney, 2000). Heaney's study and model have been extremely influential and have informed much of the development of collection description and associated services in the UK, including the Research Support Libraries Programme (RSLP), the UKOLN Collection Description Focus, and the JISC Information Environment. As such, the further adoption of this model for the purposes of CC-interop was

deemed, not only to be appropriate, but also wholly in the interests of augmenting Heaney's analysis and model yet further. The SCONE service schema is based on the Heaney model, the RSLP work, and the CAIRNS clump descriptions of Z39.50 catalogues.

For the purposes of deliverable B1, data elements were identified from a number of clump and collection description services. Since these services were all at differing stages of development, and used collection description for a variety of purposes and audiences, information pertaining to the data elements, including name, description and content standards, were compiled into a data dictionary to make comparison easier. The data dictionary was also used to record primary and secondary key fields. These provided indications of how records and fields are related where the source is not a relational database.

Several of the services investigated contained information about data elements used in service description. JISC Collection Description Service, for example, covers a variety of online service technologies, whilst SCONE describes only Z39.50-based catalogues. As a result, service data elements were added to the dictionary to ensure comprehensiveness, and in order to inform SCONE developments in this area. Data dictionary entries were then mapped against Heaney's set of attributes, and data definitions and content standards were considered in relation to Heaney's attribute descriptions to determine the best correspondence. The data dictionary and attribute mapping were then used to compare data elements for structure, congruence, overlap, gaps, and granularity.

The mapping is given in Appendix 1 of the B1 deliverable report, 'Extending the SCONE Collection Descriptions Database'.

Extending the SCONE collections database and database structure; Improved, standards-based, collections database for RIDING; Joint database encompassing both; all with COPAC strengths

These tasks sought to accommodate the recommendations of B1 and implement an extended database structure to SCONE, allowing for many-to-many relationships where possible (e.g. to allow a particular instance of an attribute to be applied to more than one collection, and a collection to have more than one instance of an attribute).

An investigation into suitable content standards was initiated, but following consultation with MLO, CILIPS and SLIC it became apparent that identifying appropriate terminology sets or categories for many of the new attributes was problematic. This issue was forced during implementation when it became clear that the majority of the extra attributes were not required in the existing SCONE structure. Nevertheless, consultation with the HILTII project facilitated the development of a pilot terminology set for music notation to ensure thorough testing. Future changes or developments to the said terminology set can be accommodated with minimum re-keying by utilising conventional RDBMS techniques.

The SCONE database and SCAMP updating interface were then cloned for RIDING. The cloning process involved changing references from SCONE to RIDING throughout, altering logos, amending the colour scheme, changing

copyright notices, and so forth. SCONE was cloned on three separate occasions to reflect in RIDING changes that had occurred within SCONE itself.

Collection-level descriptions used to populate the cloned database were then created from information available via the RIDING public service website and the websites of associated member libraries. All of this information could be accommodated in the extended SCONE structure.

The RIDING clone was then tested for CC-interop by RIDING staff. Queries, responses and feedback gathered from this testing was used to develop and improve context-sensitive help screens in SCAMP and inform the tutorial/guide for creating collection-level descriptions. Further comment was also sought from the RIDING cataloguers and to this end the RIDING clone was demonstrated at a specially organised seminar.

Due to concerns over sustainability of the RIDING service and the onset of several Z39.50 software upgrades at RIDING, the clone was never made publicly accessible. These developments, in turn, delayed implementation of a joint database, which was eventually abandoned when RIDING withdrew from further participation in CC-interop.

Lastly, COPAC was added to SCONE and CAIRNS. COPAC strengths were then implemented via the creation of associated collection-level descriptions applied at a corresponding level of granularity to allow COPAC to be selected as a result of landscaping through SCONE or directly as a target through SCONE via the CAIRNS framework.

Report, guidelines for best practice, and implementation strategies for cataloguing and indexing standards (B3)

A key output of the CAIRNS eLib project was an examination of interoperability problems caused by variations in cataloguing and indexing practices in participant libraries. CAIRNS identified the nature of these difficulties in detail and then agreed short-term and long-term strategies for addressing these difficulties, together with a strategy for implementing these strategies. Deliverable B3 of CC-interop sought to build on the previous work conducted by CAIRNS and Music Libraries Online (MLO), with a view to agreeing a similar set of guidelines and short and long-term strategies applicable to enhancing interoperability within all of the regional clumps and, if possible, for COPAC-enhanced clumps also. An integral component of achieving deliverable B3 depended on intellectual input from the RIDING clump. However, due to the departure of the RIDING early on in the project schedule, an alternative, and arguably superior, methodology was devised.

With assistance from WPA, WPB organised two one-day workshops, held in London and Glasgow respectively, to discuss improving interoperability in distributed and physical union catalogues by implementing strategies for cataloguing and indexing standards.

- Invitations for the London workshop were sent to the CURL cataloguers email list and members of the InforM25 consortium.
- Invitations for the Glasgow workshop were sent to members of SCURL and libraries in the Further Education sector.

A revised and more generic version of the CAIRNS cataloguing guidelines was distributed in advance to attendees who were encouraged to review these guidelines in preparation for the event. Attendees were also encouraged to bring along examples of policy and practice from their local institutions, and issues they have encountered in using union catalogues, to support, contradict, and otherwise inform the draft recommendations. Both events used a programme of a number of short presentations in the morning followed by a group discussion in the afternoon.

At each event, lunch provided an opportunity for personal discussion of points raised by the presentations, leading into the afternoon's 'Extended open discussion on future draft recommendations for cataloguing and indexing practice'. The open discussion was facilitated, with similar topics covered by each group.

The discussion sessions were tape-recorded with the permission of the participants. Each recording was used together with notes taken during the session to produce a report of the discussion, which was then distributed to the attendees for further comment and correction.

Both events were well attended and attracted representation from most HE libraries in each geographical area. The Glasgow event also enjoyed the participation of the National Library of Scotland (NLS), FE colleges, and the Mitchell Library, the city's public reference library.

Automated routines to allow the CD database to be output in various formats

Creating automated routines to allow the collection-level description database of SCONE to be output in various formats was a key concern of WPB. The formats of interest were those that other UK users of collection descriptions, including the JISC Information Environment, can harvest and use.

Since the United Kingdom lacks a single stable format for collection-level descriptions, routines for a number of outputs were developed:

- Text
- HTML table
- MARC21 display format
- RSLP schema in RDF
- IESR collection schema in XML
- Dublin Core Collection Description schema in RDF

Work also covered the HTML output used to display CLDs in the SCONE services.

The method dictated that the latest versions of the external schema be taken for analysis and mapped to the SCONE attributes. This task varied in complexity, depending the schema, and extensive clarification of the RSLP and DC CD schemas were sought from relevant parties.

Development of a suitable routine for the output of SCONE CLDs in MARC21 required further research and, as outlined in the deliverable report itself, required discussion with the NLS and the receipt of relevant internal NLS documentation. A suitable routine was developed for outputting SCONE records for NLS collections in MARC21 display format and a sample of six SCONE records were used to test the routine. The MARC21 display format records were then converted to MARC21 exchange records using a software tool. The exchange records were then sent to the NLS by email, and subsequently loaded into a test account of the NLS Voyager library management system to ensure the structure was correct and there were no machine interoperability problems.

Guidelines on implementing collection description landscaping in the for the Information Environment

Deployment of CLDs for the purposes of landscaping in the JISC Information Environment needed clarification and testing. Work conducted via B9 aimed to develop landscaping by applying methods normally associated with item level description. This entailed exposing SCONE and the associated landscaping features to critical evaluation. To this end user feedback was gathered from a variety of LIS professionals and professional organisations within the Scottish LIS community.

The following LIS professionals tested the SCONE service:

- **Bridget Towler** - RIDING
- **Stephen Winch** - The Scottish Cultural Portal, "Scotland's Culture"
- **Emma Jones** - SLAINTE
- **Penny Robertson** - Scottish Library and Information Resources (SLIR)
- **John Robertson** - MSc Placement Student at the CDLR
- **Emma McCulloch** - Member of the HILT II project team and CDLR researcher.
- **SPEIR Workshops** - An extensive range of LIS professionals were also consulted via the work of the affiliated SLIC funded SPEIR project, based at the CDLR (SPEIR, 2004). Recent workshops held by SPEIR addressed a total of 58 Scottish LIS Professionals from the NLS, HE, FE and public library sectors, as well as professionals in the National Grid for Learning (NGfL) and LearnDirect Scotland. This platform provided the opportunity for feedback from a representative sample of the Scottish LIS community.

Professional organisations consulted were:

- **The Confederation of Scottish Mini-Cooperatives** (COSMIC, 2004)
- **The Scottish Visual Arts Group** (SVAG, 2004)
- **The Distributed National Burns Project**

It is also note worthy that SLIC development of Scotland's Culture (2004) has provided SCONE with a bona fide landscaping test-bed. SCONE's

landscaping features have been incorporated as an integral component of Scotland's Culture and constitute an important search feature of the service.

Feedback from all the aforementioned sources informed the appropriate amendment of SCONE and production of the B9 deliverable report. Analysis of CLD schemas via the work of B7 also informed amendments.

3.3. WORK PACKAGE C

To carry out the user behaviour study, CERLIM were sub-contracted to run the data gathering exercises and to collate and present the findings. A project advisory group was set up having a member from each partner in addition to the staff from CERLIM involved in the study.

Data collection³ was carried out during the Autumn of 2003 by two methods, firstly by holding user sessions with one volunteer and one member of CERLIM staff and secondly, by focus groups of librarians led by two CERLIM staff. Both of these activities took place at three sites, the LSE, Leeds and Strathclyde with on-site support arranged where possible to assist in volunteer recruitment and in the running of the sessions.

Up to 12 volunteer users attended the sessions at each site. Their session began with a general introduction and completing a pre-search questionnaire to provide basic background information and to determine the experience of the volunteer with union catalogues.

Each volunteer was tasked with undertaking 4 searches on each of the 'local' clump (e.g. InforM25 at the LSE) and COPAC and no training was given. Searches used were realistic but contrived and included such examples as finding a common item in a nearby library, finding a rare or obscure item, finding the latest edition of an item and finding items with detailed and vague bibliographic details in the task sheet. The searching sessions were recorded (with permission) using screen capture software and additionally, notes were taken during the session.

The volunteer searching session was immediately followed by an interview with the CERLIM researcher. The interviews were designed to allow flexibility so that specific questions could be asked about the search session that had just been recorded. Questions were asked in such areas as search options, library selection, results, error messages and system feedback, problems, the session as a whole and the opportunity given for general comments, either about the services searched or the session. Questions were also asked about union catalogues to gauge user knowledge and perceived usefulness of the services.

Focus groups were used to investigate the views of library staff towards union catalogues. All group attendees were recruited locally and each group contained a mix of staff from such areas as user / reader / information / bibliographic services, serials and ILL. Each group was led by one CERLIM researcher whilst the other recorded the session and took notes. Conversation was allowed to flow but discussions were based around a

³ Full details of the methodology employed for data collection can be found in the CERLIM report referenced in Section 5.

semi-structured schedule to ensure the same areas were covered by each group.

Data from the searching sessions was fed into a complex spreadsheet to aid analysis. Data from the post-search interviews was selectively transcribed and all recordings from the focus groups was transcribed. Analysis of all the available data was performed manually by the CERLIM staff.

4. IMPLEMENTATION

For each of work packages A & B, the relevant partners organised tasks depending on skills of the team members and also staff availability at varying times throughout the project period. There were regular partner and steering group meetings to monitor progress, exchange ideas and to map out future tasks. These meetings generally took place on the same day to maximise efficiency in use of people's time and in an effort to reduce project non-staffing costs. The meetings also offered the opportunity for demonstrations of various project prototypes and test sites to illustrate progress and outcomes of the various tasks.

Staff on WP A had occasional meetings to brainstorm on technical issues and on occasion, these included external experts. Such meetings were less needed on WP B due to CDLR carrying out the bulk of the tasks as defined in the Project Plan.

RIDING, with the agreement of the partners, officially pulled out of the project during the summer of 2003 due to staffing changes at the host institution. This was unexpected and led to the remaining partners, with steering group approval, re-defining the tasks within work packages B & C such that the tasks were still completed.

As noted above, CERLIM were contracted to carry out the user behaviour study to actually investigate how users approached searching on union catalogues. One representative from each of the remaining project partners⁴ plus two personnel from CERLIM formed a working group for the work package under the leadership of a single Project Manager. The project partners prepared a brief for CERLIM but also encouraged advice and input from CERLIM with their extensive experience in user studies.

⁴ RIDING having opted out of the project before the WP C tasks began.

5. OUTPUTS AND RESULTS

This section begins with a summary list of the deliverable reports for the project and then presents the project findings for each of the three work packages.

Note: All project deliverable reports and documentation are available on the project web site at: <http://ccinterop.cdrl.strath.ac.uk/documents.htm>.

WP A Reports

Nicolaides, F. (June 2003). *A comparative study of the performance of COPAC and selected independent Z39.50 servers*, WP-A-01, Issue 1, London School of Economics.

Nicolaides, F. (April 2003). *Summary - A comparative study of the performance of COPAC and selected independent Z39.50 servers*, WP-A-02, Issue 1.1, London School of Economics.

Gilby, J. and Sanders, A. (July 2003). *Transforming a clump into a Z-target: a feasibility study*, WP-A-03, Issue 1.0, London School of Economics.

Cousins, S. and Sanders, A. (September 2003). *Establishment of an M25 Z39.50 target and testing using a COPAC interface*, WP-A-04, MIMAS, University of Manchester.

Cousins, S., Gilby, J. and Sanders, A. (March 2004). *Bibliographic Union Catalogue Results and Display Issues*, WP-A-05, Issue 1, London School of Economics.

WP B Reports

Dunsire, G. (2002). *Extending the SCONE collection descriptions database for CC-interop*, Centre for Digital Library Research, Glasgow.

Dunsire, G. (2004). *Collection-level descriptions in the Scottish Collections Network (SCONE): A guide and manual for maintaining the database*, Centre for Digital Library Research, Glasgow.

Dunsire, G. & Macgregor, G. (2004). *Improving interoperability in distributed and physical union catalogues through co-ordination of cataloguing and indexing policies*. Centre for Digital Library Research, Glasgow.

Dunsire, G. (2004). *Output formats for collection-level descriptions from the SCONE database*. Centre for Digital Library Research, Glasgow.

Dunsire, G. (2004). *Collection landscaping in the common information environment: a case study using the Scottish Collections Network (SCONE)*. Centre for Digital Library Research, Glasgow.

WP C Report

Booth, H. and Hartley, R.J. (2004). *User behaviour in the searching of union catalogues: an investigation for Work Package C of CC-interop*. CERLIM, Manchester Metropolitan University.

5.1. WP A

Semantic interoperability and index composition

A study was carried out on the relative system performance of COPAC and selected CURL library Z39.50 servers by a series of searches of types identified in the UKNUC study. The key findings were:

- That the various individual Z39.50 configurations permit *technical* interoperability relatively easily, but that only limited *semantic* interoperability is possible with the current system configurations. Any national system such as the IE that seeks to utilise such component services must address this issue and the Project recommends that adoption of the Bath Profile be encouraged. Changing system Bib-1 configurations to enable Bath compliance is technically feasible but may have financial implications.
- Disparate cataloguing and indexing practices are an impairment to semantic interoperability, and is especially pronounced for 'Subject' and 'General Keyword' searches. The Project recommends the adoption nationally of cataloguing and indexing guidelines in an effort to improve semantic interoperability. It may be noted that this initiative has already been taken in Scotland.
- The COPAC system exploits a feature peculiar to physical union catalogues. Entities have been catalogued to various degrees of comprehensiveness in different library catalogues. Consolidation of records in COPAC means that for any bibliographic entity, it is enabled potentially to derive index entries from records submitted by several institutions. This can enrich the indexing, which greatly improves the potential retrieval of relevant records.

Technical interoperability

The concept of Z39.50 to Z39.50 middleware was investigated and a prototype system developed. Such a system permits a distributed catalogue to be searched via Z39.50 from such end-user services as COPAC, another clump, an individual Z39.50 client and other IE services. Significant outcomes were:

- The JAFER software (as developed under JISC 5/99) meets many anticipated needs for distributed catalogue services and could feasibly be used by the current clump services. Further exploitation of JAFER is strongly recommended as a key building tool for the IE.
- It is technically possible for the end user to select all or a sub-set of available end destination Z39.50 servers within the middleware (often termed 'landscaping').
- The institutional origin of each record is included.
- Access control and authorisation (to Z39.50 servers) is controlled by the middleware and thus transparent to the end user.
- Query reconfiguration can be carried out within the middleware to ensure searches work optimally with different end destination Z39.50 servers. (It

may be noted that this functionality would not be necessary if the Bath Profile was widely adopted.)

Results processing

A comparative study was carried out on what results processing options COPAC and distributed services offered the end user and whether COPAC functionality could be applied in the distributed environment. Key findings and areas for possible future work are:

- The primary difficulty with the application of the results processing functions of record matching, de-duplication and consolidation, results ranking and sorting, is that most, if not all distributed services do not bring back complete results sets from associated Z39.50 servers to save time in presenting the end user with results.
- The COPAC on-the-fly routines could feasibly be applied to the clumps services (to give the end user consistency of functionality).
- The COPAC routines have evolved over time and could possibly benefit from being reviewed as user needs are continually changing.
- The routines used by COPAC at the point of data load, have not been tried in a distributed environment due to concerns that the time penalty for this may be too high. The application of these routines should be tried as the algorithms will greatly enhance the end user display.

Response times

An automated search was set up to send a query once an hour, 24 hours a day for nearly three months to monitor the response times of 17 library system Z39.50 servers. Notable findings were:

- Over 90% of responses were received in under one second with some systems responding in less than 0.125 seconds.
- Some servers showed a broader spread of response times and others showed clusters of fast and markedly slower response times worthy of further investigation.
- No servers showed slower response times during what would be considered periods of heavy use of the locally associate library opac (i.e. mid morning to early evening).
- The generally good performance of the servers suggests that many response time problems experienced by end users carrying out distributed searches, may be the result of non-responses by a Z39.50 server and how that situation is dealt with by the client software. These issues are worthy of further research and are something that may be controlled more by the Z39.50 to Z39.50 middleware.
- Further investigation is recommended to look at response times for different search types including complex Boolean queries. Also to ascertain variations in the time required for record transferal from Z39.50 server to client which could have positive implications with respect to results processing routines in distributed environments.

5.2. WP B

Report of collections description standards requirements

The primary output of this task was the associated project deliverable report, 'Extending the SCONE Collections Database for CC-interop'.

The report indicated a degree of consensus between services and Heaney's model, and identified areas where SCONE was required to engage in development in order to match that consensus. A consequent output was the amendment of the SCONE database and associated SCAMP updating service to accommodate additional attributes outlined in the table below. Detail of the data elements used was determined by the availability of content standards and external datasets.

The report also notes a number of issues that might affect the interoperability and retrieval of collection level information in services dealing with multiple collections:

Identifiers: Urgent consideration needs to be given to establishing a means of assigning unique, standard identifiers to Collections and Locations. Without such a mechanism, interoperability of locally-created descriptions in a global service will be severely impaired.

Name headings: Interoperability between collection and item level descriptions may be significantly improved if collection description services adopt the same forms and formats used for library catalogues. Name authority services should be extended to cover collection level names.

Personal name data elements: Data elements for personal names should be considered at a lower level of granularity than the whole name, to improve flexibility of display for searching, browsing and description.

Date range ambiguity: Care should be taken to avoid ambiguity in the four types of date range associated with a collection. Conflation of types should be explicitly stated in system and user documentation.

Physical location data elements: Data elements for the address of physical locations should be considered at a lower level of granularity than the whole address. In particular postcode, region and country should be considered as separate elements, to improve searching for geographical proximity. Consideration should be given to interoperability with geographical information systems; these are likely to provide a much richer functionality for retrieving collections and services on a geographical basis.

Service level description: Service level description below the highest levels of granularity requires further analysis, similar to Heaney's analysis for collection level description. This will be necessary to identify data elements to support task-oriented interfaces for service description services; for example, searching for services which support bibliographic record download.

More intriguingly however, deliverable B3 updated Heaney's original work, enhancing analyses of relationships between collections and identifying elements within attributes, both of which, by Heaney's own admission, required further analysis and research.

The B1 report indicated that the SCONE collections database would be extended with the following attributes:

Entity or Relationship	Attribute
Administrator: Corporate body	Logo
	Sector
Collection	Audience level
	Content: Audio-visual: Medium of performance
	Content: Cartographic: Locus
	Content: Graphic: Medium
	Content: Music notation: Notation system
	Content: Realia: Medium
	Content: Sound material: Language
	Content: Sound material: Medium of performance
	Content: Sound material: Type of sound
	Content: Text: Language
	Subject: Date coverage
Sells-To	Subject: Object
	Subject: Place coverage
	System of arrangement
	Rights

Accommodation of these attributes in the structure of the database has been completed. For further information see Appendix A.

Further details of the database structure can be found in the online SCONE data dictionary (SCONE service, 2004).

Identification and implementation of suitable content standards has taken place throughout the course of CC-interop.

Extend SCONE collections database and database structure; Improved, standards-based, collections database for RIDING; Joint database encompassing both; all with COPAC strengths

The wide scope of deliverable B2 resulted in the creation of a number of outputs.

- A cloned and fully customised rendering of the SCONE database and SCAMP updating interface was created for RIDING. As outlined in the methodology, collection-level descriptions pertaining to those collections located and held within RIDING were created and used to populate the cloned database.
- As a fully functional reproduction of SCONE, the RIDING collection description service successfully tested and proved the robustness of the SCONE model by employing it to a different information environment.
- Approximately 120 context-sensitive help screens were developed for the SCONE updating interface (SCAMP) and the RIDING updating service. These included guidelines on content standards, definitions of record fields, etc.

- The guide to the SCONE database and how it can be used for various aspects of collection-level description has been significantly expanded and now includes 15 sections. This guide is available via SCAMP and the RIDING updating service.
- A comprehensive SCONE data dictionary containing details pertaining to the table and field structure of the SCONE database was created. This is available online via the SCONE service.
- A complete guidance manual was produced for maintaining SCONE and the SCONE derived RIDING collection description service entitled, 'Collection-level descriptions in the Scottish Collections Network (SCONE): A guide and manual for maintaining the database'. This document is available via the CC-interop project web site.

Report, guidelines for best practice, and implementation strategies for cataloguing and indexing standards

The WPB report, 'Improving interoperability in distributed and physical union catalogues through co-ordination of cataloguing and indexing policies', was the primary output of this deliverable. This report identifies a number of issues pertaining to improving interoperability in distributed and physical union catalogues, specifically within the UK HE sector, but also within the National Libraries, the FE sector and large public libraries.

The WPB report, 'Improving interoperability in distributed and physical union catalogues through co-ordination of cataloguing and indexing policies', was the primary output of this deliverable. This report drew to the attention of CC-interop a plethora of issues pertaining to improving interoperability in distributed and physical union catalogues, specifically within the UK HE sector, but also within the National Libraries, the FE sector and large public libraries also. The general outcomes of these workshops can be gleaned from the associated report.

The report indicates a clear consensus among the cataloguing community for the adoption of prescriptive cataloguing guidelines by libraries contributing to union catalogues, helping local cataloguers to actively improve interoperability and raising awareness among senior management of the wider impact of local policy. There is general recognition that a union catalogue itself may become a component of a larger distributed union catalogue, and therefore development of cataloguing guidelines based on CAIRNS, as originally envisaged, would require a significant amount of work. Instead, the report proposes a number of recommendations for a wider, more active and co-ordinated approach to improving interoperability. The CAIRNS guidelines have been updated and references to the specific union catalogue removed, and may be used by other consortia as a basis for local development.

Other recommendations included:

- Consortia of libraries contributing to union catalogues should develop prescriptive guidelines covering catalogue record scope and content, which account for both local and 'global' needs. These might include

a minimum input standard for the level of cataloguing and the content of entry points or headings.

- Guidelines for improving interoperability need to be developed at national and international levels, particularly within the Anglophone community, and suitable mechanisms for doing so should be identified or created.
- Consortia should consider participation in international activities such as the Program for Collaborative Cataloguing (PCC). This would reconcile clashes between local and global name and subject headings, and ensure future interoperability with international distributed union catalogues.
- Consortia should consider developing a shared cataloguing service for digital resources, involving the creation of only one catalogue record to be used or copied by all local libraries.
- Catalogue consortia should develop mechanisms to ensure regular opportunities for cataloguers to discuss issues and review policies and practices.
- Further development work on the Bath Profile should encompass recommendations for the scope and content of specified indexes. This would give cataloguing consortia, system vendors, and Z39.50 service developers a firm base for establishing standard mappings from metadata formats such as MARC21 to Z39.50 indexes.
- Consortia using Z39.50 should consider producing guidelines on required conformance with the Bath Profile, specifying conformance areas and specific indexes and searches.
- Standard rules for index content normalization should be adopted at as wide a level as possible.
- Consortia and individual libraries should monitor the implementation of the FRBR model to plan for large-scale machine processing of catalogue data to improve interoperability.
- Individual and union catalogue services should disclose those local practices that may affect interoperability for an end-user. This might be embedded within the catalogue interface, or offered in help, orientation, or training screens.

Consortia should consider agreeing a standard set of information about each catalogue, which should be disclosed as part of the union catalogue service, allowing additional information to be disclosed on the local catalogue interface at the discretion of the library.

Automated routines to allow the CD database to be output in various formats

The primary output of deliverable B7 was the associated report, 'Output formats for collection-level descriptions from the SCONE database'. This report identified a number of ambiguities and anomalies in external schema, and in the SCONE database structure itself. Recommendations were made for resolving these incongruities.

Recommendations for SCONE included:

- Further investigation into the structure needed to accommodate categories of access restrictions commonly required by organizations sharing collection-level metadata, and development of the SCONE database structure to improve output for MARC21 and other relevant formats.
- Refinement of the SCONE database structure to better accommodate MARC21 name authority headings.
- Addition of genre headings as a terminology set to the SCONE database structure.
- Development of a method of obtaining URNs for SCONE collection-level descriptions output in RDF format.
- Action on the earlier recommendation to establish a mechanism for assigning unique standard identifiers for collections and their locations across the UK and beyond.
- Development of the routine to output SCONE descriptions in RSLP format to parse and map Dewey Decimal Classification numbers, if required.
- Continuation of monitoring and working with the RSLP schema if it is further developed.
- Development of the SCONE database to accommodate the Alternative title attribute and clarification of where title acronyms are to be stored [in RSLP and DC CD].
- Amend the SCONE education levels to encompass the UKEL set.

Extensive recommendations were also made for the JISC. These included:

- Consideration of adding an attribute for local identifiers such as shelfmarks to the RSLP schema.
- Investigation of the whole issue of collection strength and significance, and the feasibility of creating a set of comparable, coherent, and consistent indicators which can be used by SCONE and other service implementers.
- Clarification of the distinction between Physical characteristics, Size, and Format attributes used in the RSLP, DC CD, and IESR profiles.
- Clarification of the definitions and relationships of the attributes Rights and Legal status in the RSLP, DC CD, and IESR profiles, and development of the SCONE database to accommodate them if necessary.
- Clarification of the definitions and relationships of the attributes Rights and Use rights in the IESR profile, and development of the SCONE database to accommodate them if necessary.
- Clarification in the IESR profile of which type of agent, and whether collection is preferred to agent, as only one instance is allowed.

- Amendment of the IESR profile to include LCSH encoding of spatial topics.
- Clarification in the IESR profile as to restrictions on the occurrence of the usesControlledList attribute.
- Clarification of the distinction between Physical characteristics and Size attributes in the DC CD profile, and development of the SCONE database to accommodate them if necessary.
- Clarification of the definitions and relationships of the attributes Rights and Legal status in the RSLP, DC CD, and IESR profiles, and development of the SCONE database to accommodate them if necessary.
- Clarification in the DC CD profile as to whether Agents as subjects are scoped in the Subject attribute.

Guidelines on implementing collection description landscaping in the for the Information Environment

Clarifying the deployment of CLDs for the purposes of landscaping in the JISC Information Environment was, as outlined in the methodologies section, a key concern of WPB. In particular, the intelligent selection of targets in clumps utilising CLDs based on dynamic landscaping wa.

There were two major outputs for deliverable B9.

- The methodology for B9 informed the development and consequent enhanced functionality of SCONE. A noteworthy output of B9 is therefore the establishment of an enhanced landscaping and collection description service for the Scottish Common Information Environment. This output, in turn, informs associated knowledge and developments in the wider UK JISC IE.
- Guidelines and report on the application of CLDs for landscaping entitled, 'Collection landscaping in the common information environment: a case study using the Scottish Collections Network (SCONE)' was produced.

Whilst the report for B9 contains no specific recommendations, there are several issues worth noting.

- WPB tested and instantiated a 'm2m' interface between SCONE and CAIRNS enabling landscapes in SCONE to be transmitted as targets in CAIRNS.
- As outlined in the methodology for B2, COPAC was added to SCONE and CAIRNS and COPAC strengths were then implemented via the creation of associated CLDs. This work was further enhanced via the work in B9.

Several libraries in CAIRNS also regularly contribute to COPAC. These institutions include the University of Edinburgh, University of Glasgow, University of Aberdeen and the National Library of Scotland. With COPAC as a target in CAIRNS as well as the local Z servers of these libraries, the environment contains multiple sources for item

discovery within the same collection. Functional granularity allows the creation of a collection-level description for the distributed super-collection corresponding to COPAC. This description is linked hierarchically with collection-level descriptions for the COPAC member libraries.

SCONE descriptions inherit data from super-collections to avoid duplication. For example access conditions often apply to all sub-collections of a particular collection. In particular, information about online catalogues and Z targets is processed in this way. B9 developed a mechanism for inheriting and displaying links only to the nearest, or most co-extensive, of these finding aids to the collection-level description being displayed in SCONE.

This is generally advantageous for two reasons:

- The higher levels of collection metadata aggregation are not disclosed, simplifying the landscape for the user. The user can traverse the collection hierarchy to expose higher levels.
- By selecting the most co-extensive Z target, the dynamic clump generated from a collection landscape is less likely to retrieve unwanted items.

A specific advantage of this approach is that switching off a target in CAIRNS because of temporary unavailability results in the nearest active target relative to a specific collection being found from higher levels of the hierarchy. So if the Edinburgh University Library Z target becomes unavailable, the COPAC target is automatically displayed instead, effectively becoming a back-up finding-aid in the environment. Where such a higher-level aggregated Z target exists, the environment is less subject to disruption by local system conditions.

- In conjunction with the work undertaken for deliverable B2 and to further test the SCONE landscaping model, collections held in the M25 Consortium were searched to identify any with significant Scottish content. Any such collections found were to be added to SCONE.

The MacColl and Seeger Collection at Goldsmith College Library, London, was the only collection identified. Since this collection constitutes a sub-collection of Goldsmith College Library, which in turn is a sub-collection of the InforM25 collection (defined using functional granularity), both super-collections were also added to SCONE. Of course unlike COPAC, InforM25 is not Z39.50 enabled and therefore cannot be added to CAIRNS to facilitate dynamic landscaping of the clump via SCONE. Still, further development of JAFER, as undertaken by WPA and outlined in section 5.1, offers the possibility of hyper-clumping, with InforM25 gaining this functionality relative to CAIRNS. Nonetheless, the MacColl collection hierarchy allows SCONE users to expose the InforM25 client as an online catalogue if they choose to broaden the landscape.

The search for Scottish collections in the London area was not exhaustive, and relied on minimal collection descriptions found on M25 consortium member websites. There are also notable examples of large collections of

Scottish interest held outside Scotland, for example in Canada, in research and cultural institutions. However, work with SCONE and anecdotal evidence suggests that there are not many collections on Scottish topics held outside its borders. This may well be the case for other regions with strong local cultures.

For this and other reasons the B9 report suggests that a future UK National Union Catalogue may best be developed as a hybrid distributed union catalogue composed of smaller physical, harvested or distributed union catalogues with a regional focus. B9 shows that collection-level description can be used successfully to present users with a coherent and consistent information environment.

5.3. WP C

Key findings from the user behaviour study are best summarised by the following extract from the CERLIM report:⁵

“A number of lessons can be learned from the behaviour of our searchers and the comments of both searchers and librarians. On the basis of this admittedly small sample of users, it seems that union catalogue providers should

- provide clear on-screen guidance that the system is processing
- review terminology on screen to remove some of the jargon which has confused users [...]
- consider the search facilities which are offered to searchers in light of the comments [...]
- note that searchers have a wide variety of approaches to searching; whilst there was a slight majority in favour of starting with a broad search and narrowing the search down, there was also a considerable number who took the reverse approach
- note that searchers use a wide variety of data elements in their searches, with some strongly favouring author and others strongly favouring title approaches. Perhaps the most surprising finding was the frequency with which ISSN was used in searches for journals
- note that searchers wish to be able to have facilities to refine an initial search
- note the high impact that the web and web tools such as Google and Amazon are having on the expectations of searchers
- note the number of input errors and consider the use of a spell checker at data input
- note that there were requests for greater information about the content of books in the records presented
- note that there was a clear requirement for access to availability data, not just holdings data, on the part of a number of searchers

⁵ Taken from section 6, Conclusions and Recommendations.

- note that each of their services have features that met with end-user approval and therefore that they should continue to collaborate and share experience to enhance all the services
- note that there is concern amongst some librarians about the currency of the COPAC database
- note that there is concern amongst librarians and a smaller number of users about the reliability of search output from current virtual union catalogues.”

6. OUTCOMES

WP C findings will impact on the user interfaces of the services represented by the project partners, namely CAIRNS, COPAC and InforM25. Additionally, the cataloguing workshops under WP B will improve awareness of global issues.

WP B recommendations will inform the development and refinement of collection-level description services and standards.

WP B has improved the functionality of SCONE and CAIRNS, and developed additional tools for their maintenance.

The work carried out in WP C was in direct contact with users through the 1:1 user sessions/interviews and the librarian focus groups.

The outcomes of the project will form building blocks for the IE. Of particular significance is a paper on "Clumps and collection description in the Information Environment in the UK, with particular reference to Scotland", by Gordon Dunsire and George Macgregor, published in Program (Vol. 37, No. 4; 2003)

7. CONCLUSIONS

As far as is known, CC-interop was the first project of its kind anywhere in the world and still is. Its basic aim was to test the feasibility of cross-searching between physical and virtual union catalogues, using COPAC and the three functioning "clumps" or virtual union catalogues (CAIRNS, InforM25, and RIDING), all funded or part-funded by JISC in recent years. The key issues investigated were technical interoperability of catalogues, use of collection level descriptions to search union catalogues dynamically, quality of standards in cataloguing and indexing practices, and usability of union catalogues for real users. The conclusions of the project were expected to, and indeed do, contribute to the development of the JISC information environment and to the ongoing debate as to the feasibility and desirability of creating a national UK catalogue. They also inhabit the territory of collection level descriptions (CLDs) and the wider services of JISC's Information Environment Services Registry (IESR). The results of this project will also have applicability for the common information environment, particularly through the landscaping work done via SCONE/CAIRNS. This work is relevant not just to HE and not just to digital materials, but encompasses other sectors and domains and caters for print resources as well.

Like most projects, CC-interop can be described as a voyage of discovery, and like most pieces of research it has raised as many questions as it has answered. But two aspects of this project are, we believe, almost unique:

- It is the only known investigation of its kind in terms of interoperability of physical and virtual union catalogues
- It is believed to be the first project of its kind which has investigated user behaviour through observation of end-user catalogue searching rather than simply through user questionnaires

8. IMPLICATIONS AND RECOMMENDATIONS

The findings of CC-interop have the potential to prove very useful in the building of JISC's Information Environment. The vision for the IE could be described as a kind of "academic google". OPACs and union catalogues were not designed for google-type searching, but perhaps the CC-interop work can be seen as a step towards the creation of such a thing. JCIE may wish to consider whether further work should be undertaken to keep the momentum going and push the vision further. In particular, CC-interop's focus on user behaviour can be seen as an important stepping-stone for the Information Environment.

The following is a thematically organised series of suggestions on how the research could be taken forward:

System performance when inter-linking COPAC and the Z39.50 clumps

- Addressing the issue of semantic interoperability through extension of the Bath Profile work (the subject of a small but recent JISC-funded research project) to cover a wider scope and content of Z39.50 indexes.
- Analysis, modelling and implementation of attributes to provide the structured information needed in cataloguing and indexing practices to improve landscaping functionality for collections and IESR services, perhaps involving participation in the international Program for Collaborative Cataloging (PCC).

Creating dynamic landscaping or through CLDs

- Work to create an operational environment for exchanging, recycling, and repurposing collection and service metadata encompassing the full common information environment (libraries, museums, archives) to avoid duplication of effort.
- Development of a "hyper-clump" of current virtual union catalogues to build on the CC-interop discovery that clumps can become single targets in larger hyper-clumps, to ensure that collection description can produce effective landscapes in a complex environment and to improve metadata interoperability at a national and even international level.

Searching a distributed (virtual) catalogue or clump via Z39.50

- Further development and exploitation of JAFER as a key tool in building the information environment.
- Testing of interoperability between Z39.50 distributed union catalogues and other cross-searching analytical finding tools which are using new non-Z39.50 technologies.

Comparing results processing between COPAC and clumps

- Updating and testing of the COPAC on-the-fly routines in a distributed (clump) environment to see whether results can still be produced without undue time penalties
- Further investigate response times, broken down step by step, for different types of search including complex Boolean queries

- Test the ability of the middleware further in the context of the variations currently required for record transfer from Z39.50 server to client

User behaviour when cross-searching catalogues

- Carry out comparative case studies to ascertain time and cost in users searching catalogues sequentially in a non-interoperable environment as opposed to cross-searching in CC-interop fashion
- Include in the case studies an ergonomic element, for example testing how long online transactions take such as entering a query or selecting a target, perhaps in comparison to similar searches on Google or Amazon
- Test whether identical queries bring different results when single OPACs are searched
- Investigate the possibility of offering user choice, e.g. speedy search versus more thorough search, which will involve the system knowing the operational characteristics of the targets
- Further examination the wider expectations of users in the context of famous search engines like Google and investigation as to whether it is feasible to offer such enhanced functionality through cross-searching of union catalogue
- A look at user attitudes towards viewing quick results first while the rest are being retrieved (e.g. Ex Libris' Metalib does this)
- Examination of cross-searching behaviour across a wider selection of users, including FE
- Work with librarians to explore further their attitude towards the trustworthiness of union catalogues

Finally, there were two clear recommendations stated in the CERLIM WP C report:⁶

1. That there should be continued collaboration between the CC-interop partners in the development of their various union catalogues so that experience and best practice can be used to the advantage of all.
2. That, alongside continuing development of union catalogues, there is a larger study to consider what type of union catalogue(s) in terms of coverage and search facilities best meet the needs of the HE community.

⁶ Ibid.

9. REFERENCES

All project reports including the Project Plan and all deliverables are available on the CC-Interop Web site at

<http://ccinterop.cdrl.strath.ac.uk/documents.htm>

References used in the writing of this report are listed below.

COSMIC. (2004). Confederation of Scottish Mini-Cooperatives. Available: <http://cosmic.cdrl.strath.ac.uk/> (Checked 13 May 2004)

Dunsire, G. & Macgregor, G. (2003). Clumps and collections description in the information environment in the UK, with particular reference to Scotland, *Program: electronic library and information systems*, Vol.37, No.4, pp.218-225.

Heaney, M. (2000). *An Analytical Model of Collections and their Catalogues*. UKOLN, Bath. Available: <http://www.ukoln.ac.uk/metadata/rsip/model/amcc-v31.pdf> (Checked 15 April 2004)

IFLA (1998), *Functional Requirements of Bibliographic Records: final report*, K. G. Saur, München. Available: <http://www.ifla.org/VII/s13/frbr/frbr.pdf> (Checked 15 April 2004)

Nicolaides, F. (2003). A Comparative Study of COPAC and Selected Independent Z39.50 Servers - Report. M25 Systems Team, London. Available:

http://ccinterop.cdrl.strath.ac.uk/documents/WPA_server_tests_issue1.pdf

(Checked 15 May 2004)

SCONE service (2004). SCONE data dictionary. Available:

<http://scone.strath.ac.uk/service/sconeservice/ddtbsel.cfm> (Checked 10 April 2004)

Scotland's Culture. (2004). *Scotland's Culture*. Available:

<http://www.scotlandsculture.org/> (Checked 25 May 2004)

SPEIR. (2004). Scottish Portals for Education, Information and Research.

Available: <http://speir.cdrl.strath.ac.uk/> (Checked 13 May 2004)

Stubley, P, Bull, R, Kidd, A, (2001). Feasibility Study for a National union Catalogue, Final Report.

SVAG. (2004). Scottish Visual Arts Group. Available:

<http://scurl.ac.uk/about/svag.html> (Checked 13 May 2004)

Web sites

CAIRNS; <http://cairns.lib.strath.ac.uk/>

COPAC; <http://www.copac.ac.uk>

COPAC V3; <http://www.copac.ac.uk/msgw/wzgw>

CC-Interop: COPAC/Clumps Continuing Technical Cooperation Project;

<http://ccinterop.cdrl.strath.ac.uk/>

InforM25; <http://www.m25lib.ac.uk>

JAFER Toolkit Project; <http://www.lib.ox.ac.uk/jafer/>.

Joint Information Systems Committee; <http://www.jisc.ac.uk/>.

RIDING; <http://www.riding.ac.uk>

SCONE; <http://scone.strath.ac.uk/service/>

UKNUC: A National Union Catalogue for the UK?;
<http://www.uknuc.shef.ac.uk/nuhome2.html>.

The Bath Profile, Release 2.0;
<http://www.collectionscanada.ca/bath/tp-bath2-e.htm>

APPENDIX A

Extensions to the SCONE database structure for deliverable B2.

Administrator: Corporate body. Logo

Table	AgentCorp
Fields	LogoURL
Content	URL or URI.
Notes	Only small graphics are recorded to avoid interference with the SCONE service interface display.

Administrator: Corporate body. Sector

Table	LibrarySector
Fields	Code, Name, Notes
Content	Values suggested by the Scottish Library and Information Council and Chartered Institute of Library and Information Professionals in Scotland.
Notes	Values require extension to cover archive and museum domains.

Table	HasLibrarySector
Fields	AgentID, LibrarySectorID, Functions
Content	Secondary keys. Functions can be either or both of Administration or Service delivery.
Notes	Linking table providing many-to-many relationship.

Collection. Audience level

Table	MEGLevel
Fields	ListLabel, MEGLabel, Rank, Notes
Content	Educational levels defined by the Metadata in Education Group.
Notes	Only the six values at the highest level of granularity are used.

Table	HasMEGLevel
Fields	CollectionID, MEGLevelID
Content	Secondary keys.
Notes	Linking table providing many-to-many relationship.

**Collection. Content: Audio-visual: Medium of performance/Collection.
Content: Sound material: Medium of performance**

Table	PerformMedium
Fields	Code, Medium, Notes
Content	No suitable standards have been identified.
Notes	Not currently applied to any SCONE CLD.

Table	HasPerformMediumAV
Fields	CollectionID, PerformMediumID
Content	Secondary keys.
Notes	Linking table providing many-to-many relationship.

Table	HasPerformMediumSound
Fields	CollectionID, PerformMedium ID
Content	Secondary keys.
Notes	Linking table providing many-to-many relationship.

Collection. Content: Cartographic: Locus

This attribute has not been specifically added to the database structure. There are a number of different methods for describing cartographic locus requiring separate data structures. Verbal headings can be accommodated in subject topical headings.

Content. Content: Graphic: Medium

Table	GraphicMedium
Fields	Code, Medium, Notes
Content	No suitable standards have been identified.
Notes	Not currently applied to any SCONE CLD.

Table	HasGraphicMedium
Fields	CollectionID, GraphicMediumID
Content	Secondary keys.
Notes	Linking table providing many-to-many relationship.

Collection. Content: Music notation: Notation system

Table	MusicNotation
Fields	Code, NotationSystem, Notes
Content	In-house values identified from Dewey Decimal Classification Schedules and Library of Congress Subject Headings.
Notes	Not currently applied to any SCONE CLD.

Table	HasMusicNotation
Fields	CollectionID, MusicNotationID
Content	Secondary keys.
Notes	Linking table providing many-to-many relationship.

Content. Content: Realia: Medium

Table	RealiaMedium
Fields	Code, Medium, Notes
Content	No suitable standards have been identified.
Notes	Not currently applied to any SCONE CLD.

Table	HasRealiaMedium
Fields	CollectionID, RealiaMediumID
Content	Secondary keys.
Notes	Linking table providing many-to-many relationship.

Content. Sound material: Language/Content. Text: Language

Table	Language
Fields	Code, Language, Notes
Content	ISO 639-2 three-letter codes and language names.
Notes	

Table	HasLanguageSound
Fields	CollectionID, LanguageSoundID
Content	Secondary keys.
Notes	Linking table providing many-to-many relationship.

Table	HasLanguageText
Fields	CollectionID, LanguageTextID
Content	Secondary keys.
Notes	Linking table providing many-to-many relationship.

Content. Sound material: Type of sound

Table	SoundType
Fields	Code, SoundType, Notes
Content	No suitable standards have been identified.
Notes	Not currently applied to any SCONE CLD.

Table	HasSoundType
Fields	CollectionID, SoundTypeID
Content	Secondary keys.
Notes	Linking table providing many-to-many relationship.

Collection. Subject: Date coverage

This attribute was accommodated by adding a value for “subject” to the table DateRangeType, which is already used to store information about date ranges of collection accumulation and production.

Collection. Subject: Object

This attribute is not specifically accommodated in the SCONE database. Instead, values from subject topical headings are used.

Collection. Subject: Place coverage

This attribute is not specifically accommodated in the SCONE database. Instead, values from subject topical headings are used.

Collection. System of arrangement

Table	Collection
Fields	Arrangement system
Content	Free-text note.
Notes	

Sells-to. Rights

Table	SellsTo
Fields	SaleRights
Content	Free-text note.
Notes	

APPENDIX B

Abbreviations & Acronyms

AACR2	Anglo-American Cataloguing Rules, 2 nd edition
APDU	Application Protocol Data Unit
BL	British Library
CAIRNS	Co-operative Academic Information Retrieval Network for Scotland
CCIWG	CAIRNS Cataloguing and Indexing Working Group
CDLR	Centre for Digital Library Research
CERLIM	Centre for Research in Library and Information Management
CILIP	Chartered Institute of Library and Information Professionals
CILIPS	Chartered Institute of Library and Information Professionals in Scotland
CLD	Collection-level description
COPAC	Confederation of University and Research Libraries OPAC
COSMIC	Confederation of Scottish Mini Cooperatives
CURL	Consortium of University Research Libraries
DC CD	Dublin Core Collection Description
DDC	Dewey Decimal Classification
eLib	Electronic Libraries Programme
FRBR	Functional Requirements for Bibliographic Records
HILT	High Level Thesaurus
IE	JISC Information Environment
ISO	International Organization for Standardization
IESR	Information Environment Service Registry
InforM25	M25 Consortium of Higher Education Libraries Clump
JAFER	Java Access For Electronic Resources
JISC	Joint Information Systems Committee
LCNAF	Library of Congress Name Authority File
LCSH	Library of Congress Subject Headings
LSE	London School of Economics
M25	M25 Consortium of Higher Education Libraries
MeSH	Medical Subject Headings
MIMAS	Manchester Information and Associated Services
MLO	Music Libraries Online

NACO	Name Authority Cooperative Program
NLS	National Library of Scotland
OCLC	Online Computer Library Center
PCC	Program for Collaborative Cataloging
RDF	Resource Description Framework
RLN	Research Libraries Network
RSLG	Research Support Libraries Group
RSLP	Research Support Libraries Programme
SACO	Subject Authority Cooperative Program
SCAMP	Scottish Collections Access Management Portal
SCONE	Scottish Collections Network
SCURL	Scottish Confederation of University and Research Libraries
SHEFC	Scottish Higher Education Funding Council
SLIC	Scottish Library and Information Council
SUNCAT	Serials Union Catalogue
SVAG	Scottish Visual Arts Group
UKNUC	UK National Union Catalogue
UKOLN	UK Office for Library and Information Networking
URN	Uniform Resource Name
XML	Extensible Markup Language
XSLT	The XSL Transformation Language used for transforming XML
Z39.50	Network protocol that facilitates searching of remote heterogeneous databases and retrieval of data, usually via one user interface