Copac Collection Management Tools Project Interim Report

Interim Report – 31th May 2011

Partners: RLUK, Mimas, the White Rose Consortium
Funded by the JISC

Table of Contents

1. Description and Overview 2
2. Project Aims 3
3. Current Progress 3
3.1 Project organisation 3
3.2 Technical Development 3
3.3 Testing Framework 4
3.4 Summary of test reports 6
4. Use Case development and Benefits 12
Use case 1: Identifying last copies among titles considered for withdrawal 13
Use case 2: Identifying collection strengths 14
Use case 3: Deciding whether to conserve a book 15
Use case 4: Reviewing a collection at the shelves 15
5. On-going Activity 16
6. Forward Look 17
6.1 Interface 17
Design 17
Personalisation 17
User support 17
Wider value 17
6.2 Search expansion & post-search de-duplication 18
On-the-fly de-duplication 18
Search expansion 18
6.3 Retention status 18
6.4 Collaboration support 18
Appendix A - Project Gantt chart 19
Appendix B – The Phase 1 Testing Plan 20
Appendix C – Copac Collection Management Tools Project Risk Table 21

1. Description and Overview

This seven month project, starting in January 2011 will provide a live beta version of a web based collection management support service, using Copac data, and will incorporate a variety of tools to support library staff in collection development. The interim report logs the progress of the project up until the end May 2011; details the iterative development and testing undertaken; describes the collective thinking around benefits and use case examples of the CCM Tool; and looks forward to next steps within and beyond the current project timeline.

The project partners: RLUK; Mimas; and the White Rose Consortium (the universities of Leeds, Sheffield and York) are working together to develop a prototype that will inform and support collections managers in the decisions they need to make around monograph retention and disposal and collection development. The potential to produce a tool that will make a real difference to helping universities make informed decisions particularly at a time of financial constraint has resulted in a very enthused and committed partnership. The project team firmly believes that the CCM Tool development has the capacity to further the possibility of a national monographs collection, safeguarding long term access to materials for researchers, at the same time as facilitating local decisions that will save money and resource longer term.

The project uses as a foundation for its tool development an existing extensive database, a resource invested in by RLUK and the JISC over many years, the Copac database. Ideally suited to this intelligent reuse of data Copac has c.36 million records, representing the merged holdings of members of the Research Libraries UK (RLUK). This includes the catalogues of the British Library, the National Library of Scotland, and the National Library of Wales / Llyfrgell Genedlaethol Cymru. Increasing numbers of specialist libraries with collections of national research interest, as well as records for specialist collections held in UK academic libraries are also included.

The CCM Tool uses a variety of means to identify in which locations a particular item or batches of items exist. Data visualisation is being developed to provide differing views of the results for example, map views to assess quickly where items are held across the country, and graphs to indicate how many items searched for exist within specific libraries.

The first part of the project developed the technical interface to view the data in these contexts. The phase that spanned April through May has focused on iterative testing of the interface by the members of the White Rose Consortium with the findings of the testing used to inform revisions to functionality. The final phase ending in July will develop and test the interface further together with a batch facility. The project will keep in focus the long-term goal, that with appropriate funding and support, the tools will be developed into a live service, trialled by RLUK members and then made available to the broader community.

The project has been funded as part of the broader JISC & RLUK RDTF/Discovery initiative. By working with data already created and invested in by the community, the CCM tool demonstrates how data can be made to 'work harder' to deliver unrealised benefits and drive efficiencies. The project specifically addresses one of the key aims of the RDTF Vision: Support collection management processes and reduce duplication of effort. (See: http://discovery.ac.uk/files/pdf/jisc-rluk-vision-final-june2010.pdf)

2. Project Aims

The overarching aim of the project is to develop and test a service that will enable improved decision making regarding the retention, disposal, and redistribution of materials. The service will provide evidence of the wider availability of individual materials and/or collections when discussing the disposal of materials with academic staff within an institution. This project will help RLUK, JISC, and the WRC achieve the longer term aim of developing the technical framework required to support a more proactive and cohesive approach to collection management at a national level. In addition, by building the service on top of Copac data, the project contributes significantly to furthering the work of the JISC & RLUK Resource Discovery Taskforce, which aims to explore how data can be opened up and made to 'work harder.'

3. Current Progress

The project has progressed to schedule. (See Appendices A & B - Project Schedule, and Gantt Chart.) A proposal was submitted to the JISC in early February. In anticipation of approval some of the technical work began in January in line with the dates in the proposal and subsequent schedule. The project has taken an "agile" approach to development combining an iterative and incremental testing and feedback loop. Currently the technical development has been about functionality and workflow and not the aesthetics of the interface. This would be something to address for the future.

3.1 Project organisation

Mimas is the lead for this project and has put in place an agreed Partner Agreement. The University of Leeds is the lead contact for the WRC whose principal role is the testing and feedback of the CCM Tool, looking at how it integrates into workflows and developing use cases. RLUK is leading on external communications and outreach.

All project documents are stored in a shared file space giving access to all project partners. Internal project communications is achieved through a combination of email, telephone, and face-2-face meetings. Partner project meetings took place on 1st April and 6th May with a further meeting scheduled for the 8th July.[1] Both the RLUK news website and Twitter feeds, and the Copac Blog have been used to inform a wider audience of the projects' existence and progress. Members of the project attended the "Strategic Management of Monographs Discussion" on 17th March.

3.2 Technical Development

The initial phase was in line with the development described in the proposal and made available to demonstrate at the project meeting on the 1st April, and to test thereafter. It incorporated:

- The creation of a pilot web-interface.
- Access via IP address checking.
- Facility to search for a set of records by entering a comma delimited set of local record identifiers or standard record numbers via a text box. The initial limit on the number of records in a set (~100).
- Result set display including holding libraries.
- Record export in MODS format.
- Option to view a map of the results to see where the documents are held.
- Option to see a graph showing the number of records held by each library.

[1] Minutes of meetings are available on request.

- Option to export visualisation data as comma delimited file.

Note that the MARC21 export originally recorded for this phase of the project required some further discussion over library requirements, and now agreed, the implementation has been moved into the Phase 3 development plan.

As the University of York is not a fully participating library in Copac steps were taken at an early stage to load their data into the database to ensure that would be able to fully engage in the testing activity. The data load was fully completed on the 9th May.

Following initial feedback the following modifications were implemented and released for testing:

- The addition of an the RLUK number search to facilitate searching for other libraries' records;
- The visualisation graph brought to the top of the screen to reflect expressed priorities and the map put further down in the display;
- A holding library "count" added to the brief record display screen.

Phase 2 of the project schedule incorporated an intensive period of testing conducted by the WRC who reported at the project meeting on the 6th May. This feedback has consequently informed on-going batch interface development and the refining of the web interface functionality. It has also informed priorities for the pilot beyond July 2011. (See the "Next Steps" section below.)

The Phase 3 activity of the project schedule, initiated in April and extending to the beginning of July includes technical development on the batch search, the subject search, and search limit. The batch search has already been identified as a desirable option providing a facility to upload a file of standard numbers for batch searching, with RSS update on result availability. This was demonstrated at the 6th May meeting, and release on the27th May. Detail of what is currently incorporated can be found in section 5 below.

The testing reports and feedback on Phase 3 will be discussed at the scheduled full project meeting on the 8th July.

3.3 Testing Framework

The project has agreed core criteria for the testing framework and areas for development. This work has been summarised in the table below. Potential methods for testing and assessment, forming the basis for the Testing Plan (provided in Appendix B) have been derived from this analysis. The discussion around this area has also informed the project's Risk Assessment summarised in the table in Appendix C.

Criteria	Comment	WRC Notes	Assessment
1. Can I release this book?	*Noted risk – currency of content because of variable frequency of library updates. May be mediated by access to live circulation data.* *Noted risk – Split location & item record likely to cause false results; especially for items which initially appear to be very rare. Is suggested that accuracy will be improved in the 'new' Copac with the major rebuilding of the database, though this is unlikely to be available within the timeline of this Project.* *Noted risk – Confidence in distinguishing between different*	Does the tool allow easy-to-use, clear identification of numbers and distribution of duplicated copies of monographs held elsewhere in the UK (to support localised withdrawal of 'zero use' items with assumption that sufficient copies remain to support future research needs). Should ideally allow for comparison of all Copac member's holdings, or for sub-sets and highly customisable searches (e.g. of particular institutional collections) by end-users. Criteria 1 is arguably more likely	See Below for detail. General issues for assessment are gaining an end-users impression of the clarity of interface, availability and easy use of options (e.g. more advanced customisable searches), speed and accuracy of responses, interpretation and manipulation of varied formats of results, and the ability to download data and link these results back to local LMS for local editing and record annotation. Compare accuracy of results with those from similar search in previous WRC work. Would also seek feedback from WRC

Criteria	Comment	WRC Notes	Assessment
	editions, imprints and bindings of monographs may be of varied importance to different users in stock editing. This would be assisted by the provision of variable 'fuzzy matching' of accuracy which could be modified directly by end-users – possible approaches to be explored.	to be used by administrative/ clerical staff in editing procedures, as such tool interface and results should be particularly robust and clear.	institutions on how the Copac Tool could possibly support our users (i.e. students, researchers, academics) and what search options, and results would be particularly useful, developing some use cases. Of major importance in Criteria 1 is tying the results of the search back to the individual copy which was submitted by the user, this for marking items for potential discard or retention (of particular importance for batch search process). However, role and functioning of "fuzzy matching" will also be evaluated.
2. How does my collection compare in strength to other UK Libraries?	*Noted risk – Confidence in distinguishing between different editions, imprints and bindings of monographs may be of varied importance to different users in stock editing. This would be assisted by the provision of variable 'fuzzy matching' of accuracy which could be modified directly by end-users.*	Does the tool allow easy-to-use, clear identification of the numbers and distribution of duplicated copies of groups of monographs held elsewhere in the UK. Should ideally allow for comparison of all Copac member's holdings, or for sub-sets and highly customisable searches (e.g. of particular institutional collections) by end-users. Criteria 2 is arguably more likely to be used by specialist staff, the tool interface and results should therefore support a more flexible approach to use and offer more powerful customisation and display options.	As Above, though particular attention will be paid to users ease of interpretation of the results of 'collection strength' searches. Is the potential for confusion and/or misinterpretation of results when comparing duplication levels for a number of records across a number of institutions? Of 'potential' importance in Criteria 2 is tying the results of the search back to the individual copies which were submitted by the user, this for marking items which may be unusually rare (or common) elsewhere in the UK. The potential of different types of search and, again, the role and functioning of "fuzzy searching" to be evaluated.
Areas to look at	**Comment**	**WRC Notes**	**Assessment**
1. Interface			
Search limit – currently 100 numbers	Not very meaningful for criteria 2, but OK for 1.	Criteria 1 a) Individual items – quick search. Criteria 1 and 2 b) Small groups of items (up to maximum of c. 100) – quick search.	**Assessment for all areas noted below is primarily by end-user testing of the tool, and comparing impressions of the process –and its applications - with colleagues (and also comparing experiences between the three WRC institutions).**
Batch size limit of 400 – 4000 more useful	To be handled in batch search process.	Criteria 1 and 2 Might be useful to test larger batches at specific times?	As Above
Results display	Need to investigate the potential of the link to csv data for locations and any further developments in that area *(see also 6)* Could be useful to group/summarise by search term. Perhaps particularly important for "fuzzy searching"?	Access to brief key bibliographic details 'on screen' and online link to full record for non-batch search.	As Above
2. Map & graph			
Visualisation of results How is it useful?	Map less important for CCM criteria 2; potentially good for	Criteria 1 Main value of map may be for	Seek feedback from WRC institutions on how the Copac Tool

Areas to look at	Comment	WRC Notes	Assessment
How could it be improved?	end-users.	identifying closest 'local' duplicated copy for researcher access. <u>Criteria 2</u> Clarity of results is critical e.g. what does the map show (single copy from a batch or multiple copies at a single site). Customisable nature of search important to assist in this interpretation. ALSO how does main batch search interact with map display?	could fit into procedures for identifying collection strengths, and sorts of use this information may have on internal collection management (and on distributed 'national collection' management).
3. Can we action on conservation and preservation?	Preservation status - transaction code (583). Experience of scale of this. Identification of legal deposit items?		Seek feedback from WRC institutions on how the Copac Tool could fit into procedures for local prioritisation for conservation, and what information would be useful.
4. Workflow and use cases	Web interface useful for one-off discards and also donations. Longer term: create mobile version that could be used at the shelves with barcode scanner giving ISBN to initiate the search. Could be useful to generate a variety of use cases for different aspects of the system.	ISBN option useful, but is anticipated that many non-ISBN items (pre 1970s) will be searched, so mobile device must allow other search inputs.	Seek feedback from WRC institutions on how the Copac Tool could fit into procedures for at-the-shelf inspection and what information would be useful. Generate use cases.
5. Potential Discards			
Set your own threshold number of duplicate copies held elsewhere and be able to also set a sliding scale.		<u>Criteria 1 & 2</u> Sliding threshold re duplication of particular value to Criteria 1, but could also be useful for Criteria 2 (e.g. to identify other institutions with especially strong/weak collections matching the batch) this could form part of identification of potential CCM partners for exchanging unwanted items in a given subject area.	Seek feedback from WRC institutions on how the Copac Tool could fit into stock editing procedures, and what information would be useful.
6. Output data			
csv output: consider fields output in light of potential uses and whether could be a batch output mode too. "MARC" format for import to LMS: required fields for "MARC" export need to be agreed.	Agreed that full MARC data is not required therefore use of Copac database is appropriate.	Export options to allow 'automated' tying of results back to local items using local control/circulation numbers. Would not anticipate including all but basic bibliographic data in downloads (e.g. author/title for quick accuracy checks) e.g. full MARC download not needed.	Seek feedback from WRC institutions on how the Copac Tool results would ideally be used to add information to local LMS records, and what format and information would be useful. Consider potential for using csv output for analysis via spread sheets

3.4 Summary of test reports

Each of the WRC participating institutions, (Leeds, Sheffield and York), undertook testing of the interface based upon the agreed criteria and the testing plan. In all, five reports were produced, three from Leeds, one from Sheffield, and one from York. For the purposes of this report, the findings have been consolidated and summarised, quoting observations and impressions **directly** from the full test reports. Function and observations are based upon the criteria in the table above.

From this feedback, areas to address in the Phase 3 development have been identified together with longer term development goals. (See sections 5 and 6 below.) The fact that the test tool is a device to explore proof of concept and to specify required functionality is reflected in the commentary of the full test reports and acknowledged through the nature of the developments described in section 6. Some comparative analysis of results was attempted with varying success and revealing areas that may benefit from further testing.

Overall impression

Despite the limitations of the early stage test interface the overall reception by the libraries was very positive. They felt there were already benefits to be gained from using the interface and that it had great potential to support their current work-flows as it developed. They also recognised the new possibilities that it opened up for a more coherent and sustainable approach to collection management, with scope for embedding collaborative principles in the core activities of each library. There was therefore great enthusiasm for taking the work forward into the next stage.

Criteria 1: Can I release this book?

Method

- Submit single control number and evaluate results. Results to be compared with those from previous White Rose project.
  - Are they clear?
  - Do they allow a decision to be easily made about individual items?
  - What visualisation is helpful?
  - What further functionality would be useful?
- Submit batch of 20 control numbers and evaluate results as per previous.
- Submit single ISBN and evaluate results as per previous.
- Submit batch of 20 ISBNs for same items as above, and evaluate results as per previous.

Function	Observations
Interface	Would like to see (in later versions and later phases) more customisable options e.g. sliding bar fuzzy/accurate matching, keywords or 4-4 author/title searches, and selecting individual sites (or groups of sites) from pick-list –and ability to save common groupings- rather than all Copac members. Also would welcome ability to distinguish (filter) between 'Legal deposit' copies (held in various national libraries in perpetuity) against their 'other' more disposable copies. Would be useful to be able to filter locations – e.g. - search only WR libraries/Regional libraries. In conjunction with map would it be possible to filter by distance? Could edition and publication date be displayed? Suggestion, to have on the standard COPAC page, a button to lead to a login screen for Librarians to use, which would allow batch submission, but giving more rich search results (in effect information provided in 'standard COPAC) than is possible with current CCM.
Search & Results	
Search using single item (ISBN)	First results page is clear, summary helpful (Title, including Author) but would be useful to also include default display of publication year and edition statement to further clarify item identification. Link going straight to XML feedback not helpful (appreciate XML data display is useful but this should be an option, not default display on selecting individual item). Prefer – as on current Copac – link should be to a full record display with option for XML

Function	Observations
	download. It is useful that the tool appears to de-duplicate search results e.g. where both a hardback and paperback ISBN (for the same work) are submitted simultaneously i.e. it doesn't return the same records twice. (However, it may be some users would want to distinguish between paperback and cased for conservation reasons.) Would be useful to have link to main record display as currently shown on main Copac catalogue. Difference in meaning for 'holding details' between Standard COPAC and the CCM. The former tells you how many copies of a monograph an institution has, whereas the latter only tells you that an institution has at least one copy. A comparison of search results undertaken by the University Sheffield looked at results from using the Copac CCM tool and the WRBL CCM tool, the same four locations identified, but no British Library. The general assessment by the University of York was that the CCM tool is considerably more accurate than the WRBL tool. In one example (ISBN 9027705836), WR tool showed 5 holding libraries, in comparison with 20 showing on CCM tool. There was a wide variance between the results produced between the two tools, and was no identifiable pattern to the difference. On occasions when the results showed the same number of libraries holding a specific item, there was a variance in the actual libraries listed. Only on one occasion did the two tools produce exactly the same results.
Search using single item (Record Number)	Although a local preference the record number may not be widely known internally. The results for Record Numbers do vary when compared to same items ISBNs, presumably due to split records. ISBNs provides potential for 'more' duplication hits (and distortion of results). It is not clear what Leeds records, or indeed at other sites, are uploaded to Copac (and therefore how much reliance we can place on searches). This confidence issue would need to be addressed in any future service. Would be useful to know if the other copies were loanable or not for inter library loan. And to be able to save searches and go back to them instead of having to do a new search. A comparison of search results undertaken by the University Sheffield looked at results from using the Copac CCM tool and the WRBL CCM tool, the same four locations identified, but no British Library.
Search using multiple items (ISBNs)	Order of results in the list display difficult to understand especially due to split records by ISBNS. Some records obviously the same title was not adjacent to each other. Could results be sorted e.g. Title (and/or offer options to user re-sorting results e.g. Author, Publication year etc.) Tying back results in this form to individual items at a local institution is difficult, particularly if results are to be used for work on individual volumes e.g. stock–editing or conservation prioritisation. Results not in an easy order to follow i.e. not in ISBN order, so some identical ISBNs were not together in the list of results. Also five of the records shown were not the same ISBNs as searched for. They had similar titles to the ones searched for, but could have been other editions, and there was not enough information in the results to clarify this. A comparison of search results undertaken by the University Sheffield looked at results from using the Copac CCM tool and the WRBL CCM tool. A total of 185 locations were identified on the COPAC tool compared with 200 from the WRBL CCM results. However, this discounts the potential 54 locations that the five different ISBN but same titles items had. If they were added then a total of 239 locations would be identified compared with 200 from the WRBL CCM results. Only one out of the 20

Function	Observations
	records had the same locations for both sets of results and that was when both had no locations. Compared with the Local Record results, a total of 185 locations were identified on the COPAC tool using ISBNs, compared with 162 using local record numbers. However, this discounts the potential 54 locations that the five different ISBN but same titles items had. If they were added then a total of 216 locations would be identified compared with 162 from the local record results. Six out of the 15 records with similar ISBNs had the same locations for both sets of results. If the other five records, which had different ISBNs, were included, eight out of the 20 records would have the same locations.
Search using multiple items (Record Nos.)	Search findings exactly as for multiple ISBNs. However, the capacity for apparently higher duplication is magnified by use of ISBNs (in sample search of the same 50 items, the ISBN = 115 records, Record No = 46 Records) which may hamper interpretation, (University of Leeds). A comparison of search results undertaken by the University Sheffield looked at results using the Copac CCM tool and the WRBL CCM tool. A total of 162 locations were identified on the COPAC tool compared with 200 from the WRBL CCM results. Only one out of the 20 records had the same locations for both sets of results.
Visualisation	
Search using single item	**Map** useful for single items, could be used by staff and academics to identify closest geographic/physical copy. Lots of 'marker pins' but is not always obvious which institution they refer to. The 'pins' are clickable so you can see who they relate to, and there is a list down the right hand side of the map of all the holding institutions. Limitation of the map approach is that if there are multiple institutions in one city/location then it's difficult to see who they are, or select one of these, e.g. London. Zoom function of map (plus satellite view) could be helpful for 'travel' details. Selecting 'holding details' in the map leads back to brief details, probably better if this went to individual institutions 'Holdings Information' screen on live Copac. **Graph** less useful, though was an alternative way of displaying locations. Text on X Axis difficult to read. Code key to right of graph good, but if there are many locations (e.g. list extends beyond size of graph, bottom codes are missing or not displayed)
Search using multiple items	**Map** not especially helpful. Assume only requires a single hit (from the list of submitted ISBNs) to produce a map entry, but this doesn't indicate those sites with higher hit rates. For the future more powerful and customisable mapping (e.g. size/colour of map display for an individual location based on number of hits) useful. **Graph** - When doing a multiple submission query, the graph is useful as it allows you to see which institutions' collections are strong and weak, in the subject area under consideration. Very interesting (though much more for collection assessment (criterion 2) work rather than processing work). Unless you could provide a graph breakdown of individual items (i.e. not all items lumped together) against duplication rates elsewhere in the UK … this really approaching batch-search and results, but ability to do a quick-and-dirty search with graphical options may well inform more detailed or specialised batch searches. If the graph could display individual records on the vertical axis and if each of the rows and columns could be totalled then the graph would be very useful, as it would show how many locations each of the records was held at, and show how many records were held at each location. If this could be saved as a csv or xls file then it could be worked from e.g. rows and columns could be eliminated as needed to produce a list of

Function	Observations
	record numbers which could be identified for withdrawal.
Exporting Data	
Single items	Difficulty exporting and printing. Ability to export graph data was very useful (although needs some familiarity with manipulating files). May be useful to split map/graph options to allow local graphing and statistical analysis to be easier.
Multiple items	Difficulty exporting and printing. Ability to export graph data for multiples is extremely useful (although again needs some familiarity with manipulating files). Would be useful to tie this back to individual items at the local site, though this more in batch-search element of the tool. As above may be useful to split map/graph options to allow local graphing and statistical analysis to be easier. The 'export visualization data' button produces a text summary of the graph and map data. We are given code, name of institution, map co-ordinates of institution, no. of records, and standard number/ISBN. The ability to export the graph, at least, into Excel would have been better (and maybe save the map as a .jpeg file).

General Comment

The key elements for the next development cycle appear to be clarifying results for results of searches against multiple items, and enhancing search and display capabilities (increasing customisation of searches).

The development of a batch-interface and process is especially critical for Criterion 1 which allows searching sites to tie the results back to individual records (e.g. through local barcodes). As an example this would assist in the automated searching of item records (e.g. duplication across UK for locally held but zero-borrowed items) allowing short-listing for potential consultation in discard or relegation activity to Stores. Tying-back batch results to local records may be possible at the local level, depending on how the individual searcher submits the records, but consideration will be needed on an approach (or series of options) which individual libraries could adopt for local need, which may require Copac to offer suggestions or recommendations on how best to submit and manipulate batch searches. Further comment on this will be provided after testing a beta batch tool, scheduled in Phase 3.

With a view to informing disposal decisions, it would be necessary to be able to identify certain copies outside of Legal Deposit libraries as potential last copies which will be retained, (in agreement with the specific library.) It would be necessary for the library to mark this in a searchable manner to ensure that they do not dispose of it in error. In addition it would be beneficial if these marked items can be filtered in or out of the search through the CCM tool.

Criteria 2: How does my collection compare in strength to other UK Libraries?

Method

- Submit batch of control numbers for records which also have ISBNs and evaluate results.
  - Are they clear?
  - Do they give useful information about the relationship between the collection submitted and national holdings?
  - What visualisation is helpful?
  - What further questions arise from this analysis and how might one find answers?

- What further functionality would be useful?

- Submit batch of ISBNs for the same items and evaluate results.
  - Are there differences from the results by control number
  - Which approach gives better answers to the questions about the collections we might be trying to answer

Repeat the first test, but with control numbers of items from the 19th century.

Function	Observations
Results List	• Difficult to locate a specific item • Difficult to get full bibliographic information about any item (MODS difficult to read) • With ISBN search, records retrieved by same ISBN are at different points in list • *Suggest sort results by author/title* • *Suggest main link is to standard full Copac display, not to MODS*
Visualisation Tools	• Bar chart of number of items held by each Library very clear and helpful • Maps less useful at this stage • *Useful to be able to copy and paste chart as separate object* • *Would be helpful to have another chart showing how many items are held just by submitting library, submitting library + 1 other, + 2 others, etc.*
Export	• Tedious to copy and paste "export" into Notepad, then load into Excel. Needs more streamlined way of accessing information for more detailed local analysis and manipulation e.g. combining or comparing different record sets • *Suggest option to download csv file of the results (in which case, do we need the on-screen display?)*

Comparison of record number and ISBN searching, University of Leeds

The aim of this was to show the results of searching for the same items by record number and by ISBN for 2 sets of recent materials. Two subjects, Colour Chemistry chosen as an unusual collection, and Chemistry, although strong, was thought to be mirrored around the country. (This hypothesis was borne out by the results).

Searching by record number showed a clear difference between the 2 sets: Colour Chemistry represents a collection with relatively little overlap in other libraries (even the British Library only seemed to have 75% of the items, but Chemistry also seemed exceptionally strong. However, when ISBN searching was used, duplicate records for the same item caused considerable distortion in the results. It is assumed that using ISBNs to search for recent items will have retrieved more or less all copies recorded in Copac. Given this assumption, Leeds is shown to hold significantly more items for Colour Chemistry than any other library except the British Library, so this collection is indeed exceptional (even the legal deposit libraries only seem to manage 75% of the Leeds holdings). But the picture for Chemistry changed significantly, with Leeds ranking 4th, after the BL, NLS and TCD.

For 19th century materials, Dyeing and Chemistry books published approx. 1840-1920 were selected (the origins of the current Colour Chemistry department at the University reside in the then Department of Dyeing). A set of 50 record numbers was searched from each. The first thing to

emerge was that, for some reason, 6 of the Leeds Chemistry titles don't seem to be recorded in Copac. Turning to the visualisation results, the initial conclusion seems much as for searching for modern materials: a clear difference in the profiles: materials on Dyeing are held by few other libraries, but the Chemistry collection too seems quite rare. Given the number of British imprints within both collections, the small number of other holdings seemed surprising, so a manual search of Copac was undertaken for all 100 titles. The actual holdings revealed a very high number of duplicate records, caused not only by the vagaries and limitations arising from record loading and matching algorithms, but also from the varying methods and standards used in retrospective conversion of older materials, making record matching that much more difficult. It is interesting to note that the de-duplication algorithms used in Z39.50 access to the RLUK database do succeed in merging some of these duplicates, and it is to be hoped that the load into the new Copac database will do likewise.

General Comment

Clearly, duplicate records pose a significant obstacle to using the CCM tools for real-life decision-making. For more recent materials, this can be overcome by expanding searches to include records which include those ISBN(s) found in the original record. For older material, this isn't viable and we need to explore what other options may be open to us.

For work on assessing collection strengths, it would also be interesting to consider whether there are "typical" profiles which might demonstrate different levels of collection strength, even given the shortcomings of the record number search. Since the concept of collection strength is relative rather than absolute, this approach would seem theoretically viable. Might it even be possible in time to develop additional tools which automatically compared the profile of a given collection against a series of stored "standard" or representative profiles? However, for work focussing on stock withdrawal, confidence in the actual number of other copies held is crucial, and the current gap between holdings on the same record and total Copac holdings is a significant obstacle (albeit it errs on the side of caution, so withdrawal is definitely safe if sufficient other holdings are shown associated with the same record).

4. Use Case development and Benefits

The success of the CCM Tool is dependent upon the benefits it offers to libraries and how easily it can be integrated into existing workflows. The testing regime has tried to address this but illustrative use cases will be important to demonstrate how this tool can be used productively. At this point in the project the WRC has developed a series of four Use Case scenarios with a view to applying and testing these in live conditions once the CCM Tool is more fully developed.

Note that the use cases have been developed in light of the participating libraries collection management requirements and experience of the first CCM interface trial, alongside the discussion arising from that work. Therefore they are expressed in terms of libraries that contribute their catalogues to Copac. Whilst the library's record number may be the ideal starting point for many types of collection assessment, other search starting points are possible, and we would expect the use cases to be more widely applicable amongst libraries that are not contributors.

In addition, the use cases do not currently address the potential of a subject search starting point for collection assessment. This will be explored through provision of a subject search as a starting point for testing and discussion later in the project.

Use case 1: Identifying last copies among titles considered for withdrawal

Context

All academic libraries are under considerable pressure regarding space: there is a constant inflow of new stock, plus many other demands on the limited space available. Most therefore have in place a programme of stock editing, with selective withdrawal of items from both teaching and research collections. It is considered best practice to check whether titles being considered for withdrawal are rare, or indeed unique, within the UK ("last copies"), and if so, to either retain that item or offer it to another library for permanent retention.

Procedure

- Using its Library Management System (LMS), the Library establishes a list of candidates for withdrawal based on various criteria such as recent usage statistics, relevance to current teaching and research interests within the institution, etc.
- Once a final list has been agreed, a file is exported from the LMS, listing the record number of each title, and this is submitted to the Copac CCM tool in batch mode. The record number is used by the tool to identify each title within the Copac database and hence the number of copies of that title held by other institutions. If any institution has recorded a reservation status for their copy (e.g. "permanent retention"), then this too can be identified.
- A file is output by Copac containing, for each title, the same record number that was submitted, the number of copies held nationally, plus any data regarding preservation status.
- The Library loads this file into its LMS, matching on record number and updating the original catalogue record with the additional information supplied by Copac.
- The list of candidate titles can then be manipulated using this additional information in order to generate a definitive list of items for withdrawal. Typically, the file would be sorted by the number of copies held nationally and any title with fewer than 5 other holding institutions would either be retained or offered to other research libraries with an interest in that subject area.

Benefits

- At present, it is a very tedious, time-consuming and expensive manual process to check all withdrawals against Copac. This facility saves both time and money as the relevant information can be made available to staff on the ground with minimal effort.
- Because the overheads are drastically reduced, individual libraries are more likely to follow best practice and carry out this check for last copies. This will help to maintain the full breadth of our national research collections and avoid the loss of material which could be of value to future generations of researchers.
- The inclusion of preservation status information provides a framework which could, in the future, allow libraries to make more informed decisions about long-term retention (and the associated conservation issues).

Use case 2: Identifying collection strengths

Context

Our major research libraries face difficult decisions around prioritisation and best use of resources. Many have outstanding collections, but budgetary and space limitations, the availability of digital surrogates and the condition of the physical stock all suggest that a more targeted approach is required. Brittle paper in particular is putting whole collections at risk since so much material published between 1850 and 1980 is slowly decaying on the shelves. Even where material has been or is planned to be digitised, it is clearly important to continue to offer access to the physical originals somewhere within the UK.

Within this context, libraries need to know how their collections fit into the national picture, which collections contain a high proportion of rare or unique materials, and which ones may be of value to researchers because they bring together a body of related material not held by any other single institution. This data can then be used to inform practical decisions about all aspects of collection management.

Procedure

- The Library believes that its collection of material on travel and exploration is particularly strong and unusual within the UK. Publication dates range from 1820 to 1914. Using functionality within its LMS, it exports a list of record numbers for the 2,500 titles within this collection. For purposes of comparison and control, a second file is also generated containing the record numbers for a second collection of biographical and autobiographical material from the same period (1,500 titles). Both files are submitted to the Copac CCM tool in batch mode.
- Using the record number supplied, the tool is able to identify each title within the Copac database. From this, statistics can be generated regarding which other institutions hold these titles and how many copies of each title are recorded within Copac.
- Staff view the resulting bar charts online via the visualisation options which form part of the CCM tools. They also download the statistics into Excel for more sophisticated local analysis and future comparison with further collections.
- The bar charts show that for the travel and exploration collection:
    - the British Library is the only institution to hold more than 70% of the titles
    - 5% of the titles are uniquely held by our Library
    - A further 15% of titles are only held by 2 or 3 other institutions

  In comparison, for the biographical materials:
    - 4 other institutions hold 90% or more of the titles
    - Only 1% of titles are uniquely held by our Library
    - A further 5% of titles are only held by 2 or 3 other institutions
- From this, it is clear that the travel and exploration collection is a valuable resource which should be prioritised for conservation and should be the focus of active promotion within the wider research community.

Benefits

Without the Copac CCM tools, there is no objective way of assessing the strengths of a library's collections, either relative to each other or within the wider national context. Such analysis opens the way to evidence-based prioritisation in many areas of library activity including conservation,

digitisation and resource allocation, with consequent improvements in efficiency and return on investment. It provides a framework for local consortia such as White Rose to engage in collaborative collection development, saving money by avoiding unnecessary duplication. It also opens the genuine prospect of a more coordinated approach to the national distributed research collection with clear advantages to both the research community and the funding agencies.

Use case 3: Deciding whether to conserve a book

Context

Conservation staff are extremely stretched by the volume of material which needs attention. Furthermore, a significant proportion of the research stock is at risk from brittle paper, remedial measures are expensive and time-consuming, and it is therefore important to use resources to best effect.

Procedure

- Library staff find a book published in 1910 in a poor state of repair during routine shelving. Pages are coming loose because the paper is brittle. They pass it to the Conservation Unit.
- Staff in the Conservation Unit call up the record in the LMS. A link on the title display offers the option to launch a search for that record number in the Copac CCM interactive tool.
- From the CCM tool, three pieces of relevant data are immediately clear:
    - The bar chart reveals the item is held by 7 other institutions.
    - The map shows that the nearest holding institution is another library in the White Rose Consortium, so there is ready access to another copy
    - Both the British Library and the White Rose Consortium library have indicated a preservation status of “Permanent retention”.
- The staff member therefore concludes that repair of this copy is not a priority and it is a better use of resources to withdraw it from stock and refer any future users to the copy held at the other White Rose library.

Benefits

- Instant access to information and easier integration into standard procedures, leading to better prioritisation and use of scarce resources.
- The inclusion of preservation status information provides a framework which could, in the future, allow libraries to make more informed decisions about retention and active conservation of individual items.

Use case 4: Reviewing a collection at the shelves

Context

Assessment for conservation and/or the transfer of material to Special Collections is time-consuming and expensive, yet is imperative if material is to be preserved for use by future generations. Although such assessments are often event-led and restricted to a single physical item, there are occasions where it is important to systematically review a whole collection. For example, many nineteenth century botanical works contain plates, sometimes in colour. There is a need to review each volume on the shelf, assess and record its physical condition, and decide on what action is appropriate (e.g. active conservation treatment, digitise, transfer physical volume to closed access, leave on open shelves in various combinations).

Procedure

- Carrying a mobile device such as a tablet, the staff member reviews each item sequentially at the shelves. The catalogue record is called up in the LMS, where a link offers the option to search for that record number in the Copac CCM interactive tool's mobile interface.
- The CCM tool immediately reveals:
    - The number of copies of that title held nationally
    - A map showing the distribution of those copies round the UK
    - Whether any of those copies have a recorded preservation status (such as "permanent retention").
- The staff member now assesses the physical condition of the volume and is also able to assess the quality and distinctiveness of the plates.
- In this particular case, noting that the item is held by four other libraries, of which two indicate active conservation and permanent retention, the decision is taken to digitise the item and then return to the open shelves. This decision is recorded in the MARC record within the LMS, from where the information will subsequently be uploaded to Copac.

Benefits

- The staff member is able to inspect the volume physically while having access to the full bibliographic record and the context of national holdings for that title (including their condition/retention status).
- From this, appropriate decisions can be made with less effort and manual collation of data.

5. On-going Activity

The following assessment of activity to take forward into Phase 3 and to the end of the current project is a distillation of plans from the original proposal, feedback from the Phase 2 testing and feedback reports from the WRC, as well as associated discussion.

Planned activity: phase 3

Having agreed MARC export format requirements we will introduce an option to export records in a brief MARC format containing truncated bibliographic details with collection related information. In addition, the proposed Phase 3 technical activity covers:

- Basic batch search.
- Subject search.
- Search limit by library.

We will implement basic forms of these facilities to allow exploration and feedback.

In response to development experience and feedback on the Phase 1 interface we are revising the search process and introducing sorted results. We will also add some content to the brief record display and amend the csv export. This will be made available as part of the batch search interface release.

As time permits, a further graph will be provided that displays the number of occurrences of an individual title within a result set.

A mechanism for indicating retention status and legal deposit copies of material is under discussion as part of the current Phase 3. The outcome of this element of the project would feed into plans going forward. (See section 6.3 below.)

After further testing and feedback activity a final report will be produced, scheduled for the 31st July, 2011.

6. Forward Look

The desire is to build upon the work of the project beyond the current limits and to develop it further. The aim, as stated within the original project proposal to the JISC would be to roll out a pilot service, in the first instance to RLUK libraries to act as first movers, applying the tool within their workflows. Feedback from this service in development phase would feed into service refinement with the aim of rolling it out widely as a full service over time. The testing of the Use Case scenarios specified in section 4 above will be an important element in this activity. The CCM development will benefit from the de-duplication and other work underway as part of the Copac reengineering, but it is not dependent on this activity.

The project team have already identified the following elements to incorporate into developments beyond the current period of the project.

6.1 Interface

Design

Interface design processes to explore workflow, presentation etc. and develop an interface suitable for public service trial. This could be a variant of the forthcoming Copac interface, tailored to the requirements of collection management activity.

This would address issues around record displays, search history, navigation, and refinement of visualisation, and export options.

This could include offering 'use case' choices resulting in different search/visualisation defaults that are seen as the most relevant for a particular activity.

Personalisation

Assess what interface personalisation options users might find valuable and how these can best be integrated.

User support

Consider the benefits of developing 'Models of use' that would illustrate the potential value of different facilities depending on the user context. Similarly, illustrations of the way in which data can be exported and utilised locally might be valuable.

Wider value

Once an interface has been developed, consider assessing its potential application in other areas. For example reviewing how well it functions in discussing collection change/disposal with academic staff; and any role in other areas such as marketing the library. As the work progresses, new use cases may be defined that will identify further development areas.

6.2 Search expansion & post-search de-duplication

Search expansion and on-the-fly de-duplication are different approaches to increasing the 'fuzziness' of a result.

On-the-fly de-duplication

Copac de-duplication errs on the side of caution to reduce the likelihood of very similar, but separate, works becoming merged within the consolidated records. However, depending on the purpose for which the collection management tools are being used, it may be better to have a less exact match procedure to give a clearer view of an individual collection in relation to the holdings of other libraries.

Introducing post-search de-duplication of result sets, with a range of match options, would allow a user to decide how much potential error they are willing to accept in the context of what they wish to find out. This required discussion of the range of de-duplication options we might provide and how these might be presented to give the user an understanding of the likely impact on the results.

Search expansion

A search for record numbers or ISBNs representing a collection gives very exact results and there may be situations where a less precise search would be valuable for gaining a broader view of the place of that library's collection in a national context. So, in addition to just carrying out a search for the supplied search terms it would be possible to provide search expansion options that would take elements from the results of an initial search and carry out a further search, merging and de-duplicating the results.

For example, a record number search might be followed by a search for the author, title and date of each item derived from the retrieved records. This would need to be discussed in terms of the range of search options that seemed both valuable and practical, as well as how these might be presented to give the user an understanding of the likely impact on the results.

6.3 Retention status

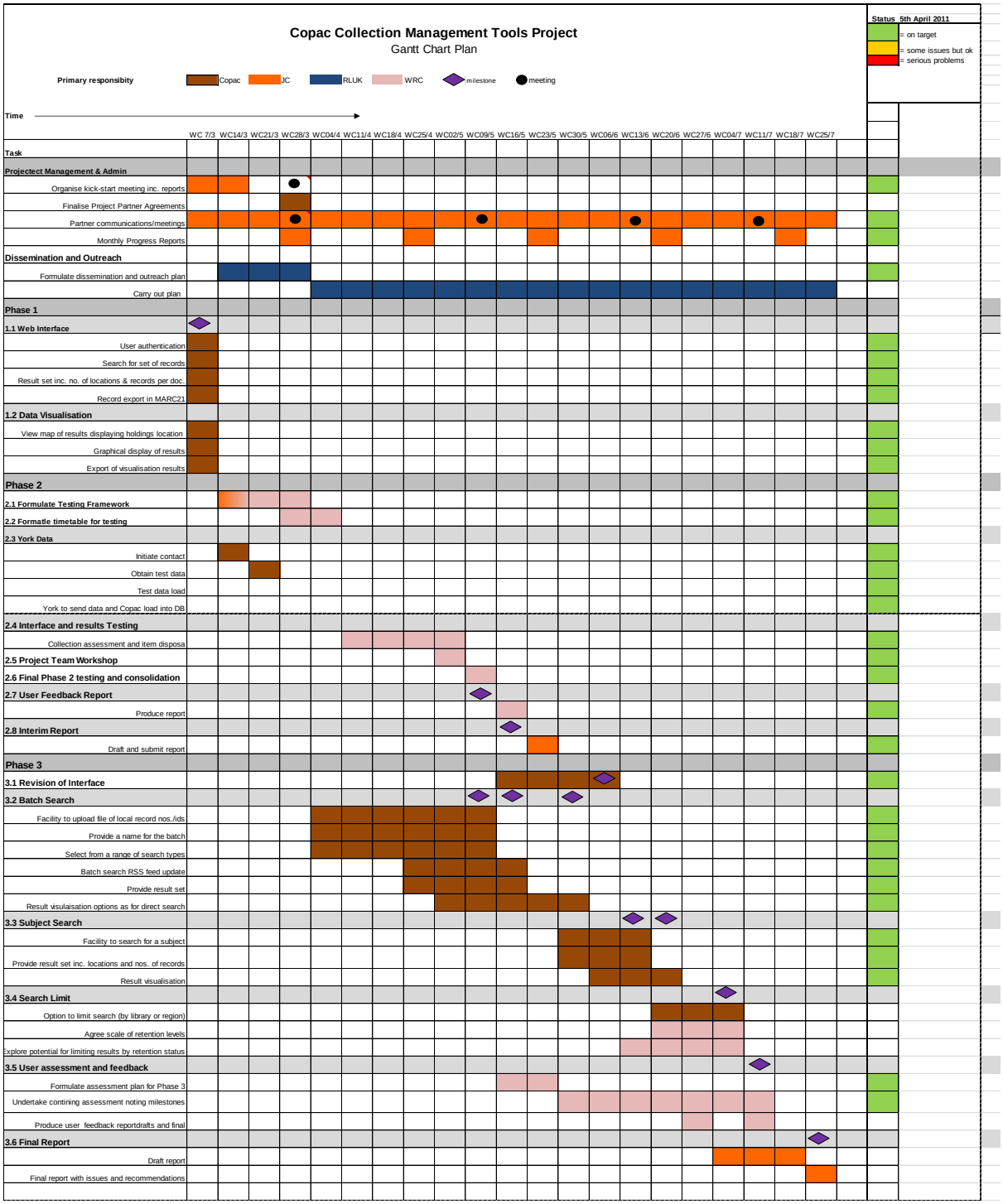
A mechanism for indicating retention status and legal deposit copies of material is under discussion as part of the current Phase 3. Once a clear identification mechanism is agreed, and ideally implemented within local catalogues, this can be taken forward. We would need to agree a clear way of presenting this in the records, as well as exploring ways of offering a search filter based on these criteria.

Within this development area it would also be valuable to explore whether there would be a reliable way of identifying materials available for interlibrary load; or material that is reference only.

6.4 Collaboration support

Given the potential benefits of the CCM tools within the context of both local and shared collection decision making it would seem valuable to explore whether there is a role for the service in supporting communication between libraries regarding collection strengths, disposal decisions, and exchange of materials.

Appendix A - Project Gantt chart



Appendix B – The Phase 1 Testing Plan

Testing actions for Phase 1

Criteria 1: Can I release this book?

- Submit single control number and evaluate results. Results to be compared with those from previous White Rose project.
  - Are they clear?
  - Do they allow a decision to be easily made about individual items?
  - What visualisation is helpful?
  - What further functionality would be useful?
- Submit batch of 20 control numbers and evaluate results as per previous.
- Submit single ISBN and evaluate results as per previous.
- Submit batch of 20 ISBNs for same items as above, and evaluate results as per previous.

Criteria 2: How does my collection compare in strength to other UK Libraries?

- Submit batch of control numbers for records which also have ISBNs and evaluate results.
  - Are they clear?
  - Do they give useful information about the relationship between the collection submitted and national holdings?
  - What visualisation is helpful?
  - What further questions arise from this analysis and how might one find answers?
  - What further functionality would be useful?
- Submit batch of ISBNs for the same items and evaluate results.
  - Are there differences from the results by control number
  - Which approach gives better answers to the questions about the collections we might be trying to answer
- Repeat the first test, but with control numbers of items from the 19th century.

Appendix C – Copac Collection Management Tools Project Risk Table

Risks (P = Probability, S = Severity, Scale 1-5)				
Risk	**P**	**S**	**PxS**	**Action to Prevent/Manage Risk**
Stakeholders				
The expectations of tool availability across the Library sector may be difficult to reconcile with the limitations of the project.	2	5	10	Include raising awareness and outreach in the projects communications plan.
Staffing				
Key staff leaving or unavailable in the short time that the project is live.	2	5	10	Redeploy existing staff; Use existing networks for formal and informal advice.
Over commitment of staff who may work on multiple activities.	5	3	15	Ensure that the project schedule aligns with staff commitment and that conflicts are brought to the attention of the Project Manager early to permit rescheduling or exploration of back filling key staff.
Organisational				
Partners are geographically dispersed and so physical meetings may be difficult to organise	5	3	15	Advance scheduling of all meetings; Use telephone conferencing & e-collaboration; Keep in regular phone and e-mail contact.
That the activity across partners gets out of sync.	2	3	6	Project Team ensure they are fully briefed and resolve issues with the Project Manager.
Communications within the consortium breakdown and partner involvement wanes. (Action to prevent:	2	3	6	Scheduling of regular updates and meetings and encouragement of active engagement, getting buy-in from partners.
Technical				
Phased development may not be delivered to schedule thus impacting on testing timetable.	3	4	12	Regular progress reports. Ensure all in project are aware of progress mapped against schedule. Build in some contingency into planning.
Not all partner requirements may be technically possible within the bounds of the project.	2	3	6	Ensure good communications and prioritise development in full consultation with partners.
Issues with currency of content because of variable frequency of library updates.	5	3	15	May be mediated by access to live circulation data.
Split location & item records are likely to cause false results; especially for items which initially appear to be very rare.	5	3	15	Is suggested that accuracy will be improved in the 'new' Copac with the major rebuilding of the database, though this is unlikely to be available within the timeline of this Project.
Confidence in distinguishing between different editions, imprints and bindings of monographs may be of varied importance to different users in stock editing.	5	3	15	This would be assisted by the provision of variable 'fuzzy matching' of accuracy which could be modified directly by end-users – possible approaches to be explored.
Legal				
Exposed data may not comply fully with existing agreements and obligations.	2	4	8	Be full aware of existing agreements and obligations and there implications.