

Copac Collection Management Tools Project

Retention and Preservation: Final Report

July 2012

Contents

Executive summary	2
Context: gathering the information	3
Information to be held and made available	5
Contributing relevant data	6
Key drivers, benefits and barriers	7
Establishing a strategic framework for collaboration	8
Standards and formats	10
Principal recommendations	12
Appendix A: Examples of 583 content, based on PDA.....	13

Executive summary

From the very beginning, one of the key drivers in the Copac Collection Management Tools Project was to find ways of improving the management of “last copies” – that is to say those monographs where there are very few copies held within the UK. In many research libraries, checking for last copies is an integral part of the stock editing process, but it is very time-consuming and so a key objective of the Project has been to find ways of automating these procedures. Detailed community consultation was undertaken, principally through two workshops and through interaction with the RLUK/BL Preservation Learning Project. It became clear that this whole area is of significant concern to research libraries. They recognise that the role and management of print collections are changing for a number of reasons including pressure on space, the availability of digital surrogates and the vulnerability of many publications to slow decay because of brittle paper. A consensus is therefore emerging that, in order to ensure ongoing access for researchers to the full range of print materials, collaborative action is essential – a National Research Collection of Monographs. Those consulted also considered that the Copac database offered an appropriate mechanism for supporting this on a distributed and community-driven basis.

These discussions also explored what is required in order to support these aspirations. The clear priority is for libraries to flag up in the Copac database those titles which they are committed to retaining in the long term. The identification of such titles could arise as a result of stock editing procedures (i.e. because they were last copies); alternatively, the commitment to retain may be implicit in the collection to which the item belongs (e.g. those acquired via legal deposit) or derived from some other process of collection analysis. Ideally, information about the preservation status of the item in question and an indication of its availability for inter-library lending would also be included, but these were very much seen as secondary to recording the commitment to retain. For such a system to be workable, it must fit easily into the workflows of participating libraries and not place heavy demands on staff time. For these reasons, the use of MARC tag 583 is recommended, in association with the PDA vocabulary.

There remain various practical issues which need to be addressed, not least a definition of “long-term retention” which is acceptable to all participants. An organisational framework is therefore required to progress this work and coordinate the establishment of the proposed National Research Collection of Monographs. A key determinant for its success will be its ability to inspire sufficient confidence within the community of research libraries. It is therefore recommended that RLUK should take the lead in driving forward this development.

Context: gathering the information

Much of the impetus for the Copac Collection Management Tools Project came from previous work undertaken by the White Rose Consortium. The libraries of the universities of Leeds, Sheffield and York had responded to pressures on space by exploring ways in which they could discard lesser used titles and yet still ensure access to those titles, either through some form of collective agreement around retention, and/or by ensuring that a sufficient number of copies were retained nationally. The British Library was also a partner in this work, and an initial phase saw the British Library accepting and accessioning any titles for which this was the last copy nationally¹. However, although all parties agreed that such work was valuable in safeguarding last copies, it became clear that this approach presented difficulties for the British Library if translated to a larger scale, and that that institution's broader strategic direction places emphasis on "connecting" as much as "collecting" (cf. the approach adopted by the UKRR for serial publications). The White Rose libraries were keen to take their work forward by exploring how holdings in Copac could be used to support local collection management decisions and avoid the loss of last copies. They also began to be concerned by the broader implications of this approach: what were the implications of holding the last copy of a title within the UK and should collection management decisions be different if a particular collection held a large number of rare or unique titles.

These concerns were already articulated during the first phase of the Copac Tools Project (January-July 2011)², but Phase 2 (October 2011 - July 2012) provided the opportunity to explore retention and preservation as a discrete work area. An initial workshop, to which the majority of Copac contributing libraries were invited, was held in December 2011 with the broad remit of discovering along what lines the Copac Tools should be developed in order to provide maximum benefit and value to the libraries present. Assistance with checking last copies emerged as one of the top concerns. There was also consensus that, for maximum value, it would be helpful for libraries to go further and indicate, where appropriate, the retention status of an item – that is to say, to indicate those titles which they intended keeping in the long term. For this to be meaningful, it implied a shared understanding of what is meant by "keeping an item in the long term" and a common method for indicating this in Copac. Parallels were drawn with the UKRR framework for ensuring access to serial titles – the data for which is stored and shared via Suncat.

Similar concerns were also coming to the fore in the RLUK/BL Preservation Learning Project (2009-2012), which was establishing a new focus and value in collection care by forging links with wider collection management concerns. A survey of preservation practice in RLUK libraries was conducted

¹ Collaborative collection management: a White Rose / British Library pilot project: executive summary. 2005. Available at: http://library.leeds.ac.uk/downloads/file/216/executive_summary

² Copac Collection Management Tools Project final report. 2011. Available at: <http://copac.ac.uk/innovations/collections-management/wp-content/uploads/2012/01/Copac-Collection-Management-Tools-Project-Final-Report.pdf>

in summer 2011 and one of the questions asked: *When deciding to withdraw items from stock do you check whether the title is held elsewhere in the UK and/or Ireland?* The answers to this question need to be treated with caution since some responses were made by the Special Collections Department of the library in question (where withdrawals are unlikely). However of the 14 libraries who considered the question applicable, 8 stated that they always checked and the remaining 6 usually or sometimes did so.

The results of this survey were used to establish the broad themes for the Preservation Learning Project's final workshop, entitled *In safe hands?* in February 2012, which had a twin focus on firstly the link between preservation policy and wider institutional strategy and secondly on this concern to avoid the loss of last copies³. One of the presentations explicitly discussed the handling of last copies while another outlined the work of the Copac Tools Project, presenting it as a possible tool to assist with preserving the National Research Collection. This topic provoked great interest among the 34 participants from 21 different RLUK libraries, and this was reflected in the lively discussion in breakout groups and in subsequent interest in the Copac Tools Project.

The next stage for the Copac Tools Project was to convene a workshop specifically on this topic, on 12th March 2012, which was attended by 20 participants from a number of different institutions and backgrounds. Particular care was taken to ensure that the major stakeholder interests were represented, including Copac contributors, the preservation community, collection managers and those with metadata expertise. RLUK, OCLC, Mimas and a number of legal deposit libraries (including the British Library) also sent representatives. The purpose of this workshop was to:

- Understand what information should be held and made available
- Explore whether libraries would be prepared to contribute the necessary data to Copac
- Identify key drivers, benefits and barriers
- Consider mechanisms for collaboration
- Initiate discussion about specifications, formats, workflows and systems required to make it work.

An initial presentation set the scene, while a second challenged the participants to find an effective response to the need of scholars for long-term access to monographic resources. The participants then split into two groups to discuss the main issues, before coming together for a plenary session which identified the common themes and issues and began to explore how the community might respond. Detailed summaries of these discussions have already been made available on the Project website⁴.

³ Workshop presentations available at: <http://www.bl.uk/blpac/safehands.html>

⁴ Available at: <http://copac.ac.uk/innovations/collections-management/2012/03/workshop-on-retention-and-preservation/>

The discussions that took place within the context of the Copac Tools and Preservation Learning projects were explicitly situated, and need to be understood within, the wider debate about collaboration between UK research libraries in preserving the National Research Collection. During the last decade, within the context of RSLP and subsequently the RIN, there has been considerable discussion and debate regarding collaborative collection management and how it might help research libraries to provide a better service to the scholarly community. The various risks to the “national collection”, both from natural processes – in particular the slow degradation of brittle paper – and from the actions of libraries, e.g. through stock editing, were identified in a number of reports. One result of this work has been the UKRR initiative which has provided a very successful model for collaborative management of print journal collections. In the past 12 months, preliminary discussions have also taken place to explore whether this could be developed to provide a similar framework for monographs. However, with more complex management issues to be resolved and little hope of significant central funding, it is proving challenging to find a way forward.

The present document recognises and draws on all this previous work in suggesting both how the Copac Tools might be developed in order to support practical collaboration in the area of retention and preservation of collections, and also how this might contribute to the realisation of that vision for the UK National Research Collection..

Information to be held and made available

In order to make an informed decision about whether to retain a print item and to base that decision at least in part upon the concept of maintaining the integrity of the National Research Collection, libraries would ideally like access to a number of pieces of information including:

- National holdings (i.e. within UK and Ireland)
- Whether any of those libraries currently holding copies of the item intend to retain it in the long term
- The physical condition of those copies
- Whether a copy would be available on ILL should it be required
- Whether an item is available electronically / in digitised format

A key concern is to ensure the accuracy and currency of the holdings information available via Copac as this underpins everything else. Concerns within this context would include:

- The existence of material which is either uncatalogued or not notified to Copac for various other reasons
- Delays in updating information on Copac

Although the national picture is important, libraries also participate in various consortia and groupings, regional or otherwise, and the nature of holdings within such a grouping will sometimes impact on decisions as well.

After this, the single, clear priority for all libraries is to understand whether a holding library is committed to holding an item in the long term. This needs to be based on a common understanding across the community of what it means to “hold an item in the long term” so that decisions can be based on trust and confidence. A second but much lower priority is given to knowing the physical condition of relevant items.

From the point of view of access to the material for end-users, the availability of digital copies as well as access via ILL are seen as relevant to local decision-making. Finding a solution to the current difficulties in establishing whether or not a digital copy exists would clearly be very helpful, as would the ability to distinguish easily and with confidence between physical and electronic holdings. What has seemed less well articulated within the community is the value of retaining one or more physical copies even when an item is available electronically, the function of those physical copies, and the mechanisms for ensuring their long-term preservation. Many librarians feel intuitively that this is essential but it would be helpful to have a clear and reasoned articulation for this requirement, together with an exploration of what it might imply for local practice.

Contributing relevant data

Despite the widespread recognition of the problem, it is also clear that few additional resources are likely to be made available locally in order to contribute the data required to support this initiative. It is therefore imperative that expectations focus on areas where libraries already hold the relevant data (albeit sometimes in a different context or format) or alternatively where any additional work can be justified as cost-effective at the local level because of consequent gains in efficiency or service delivery.

Within this context, it is clear that the ability to use the Copac Tools to support stock editing procedures is a paramount concern to the majority of libraries, something borne out by the very positive response from the Project’s Associate Libraries when they gained access to the beta version of the Tools. Significant savings in staff time can be anticipated (this is validated by the Copac Tools Project case studies) and this in turn could represent a considerable incentive to contributing data about retention status, particularly for items which have been assessed using the Tools. In this way a virtuous circle becomes possible whereby use of the Tools encourages data input and submission to Copac which in turn enhances the value to be derived from use of the Tools by other libraries.

Retention status should therefore be capable of quick and easy recording locally for such items, and able to be notified to Copac as part of the standard upload of data.

Within the community, there is also a desire to be able to understand unambiguously the retention status of material held by the six legal deposit libraries. Within the British Library, for example, there are various collections, for which that Library may adopt differing policies. It is thought that data within the catalogue record indicates the specific collection or type to which each item belongs. It would therefore help the community if permanent retention status could be logically derived from this data and expressed in a common format within Copac. There may be other Copac libraries to whose holdings a similar process could be applied – perhaps all “Special Collections material” might be amenable to such treatment?

A third and related route for adding retention status could conceivably arise from the widely reported work that Leeds has undertaken on collection categorisation. The collections strategy for that library⁵ is based on the concept that, apart from the closed access Special Collections, certain other more modern collections have a particular and enduring significance (“heritage” and “legacy” collections). And one use of the Copac Tools which Leeds has actively exploited lies in the identification of such groupings of material, based on analysis of national holdings of the constituent titles. Leeds would therefore consider flagging the items within these collections as being for permanent retention – a bulk change of this kind would be very easy and fast within their Library Management System. Other libraries at the March workshop expressed a similar view.

As already stated, it is really important for successful collaboration that retention status should be simple and fast to add locally and then upload to Copac (or for Copac to derive from existing data fields). However the system should also be extensible so that other data helpful to decision-making in other libraries (condition, creation of a surrogate, etc.) could also be included by individual libraries, should they so wish.

Key drivers, benefits and barriers

A key driver at the present time lies in the amount of stock editing that many libraries are undertaking in response to pressures on space. Faced with the need to withdraw individual items from the research collections, there is a powerful incentive to find an efficient method for identifying both rare or unique material and also those items which are widely held and so can safely be discarded. Current methods are time-consuming and cumbersome – both cost savings and greater productivity/throughput can be anticipated from use of the Copac Tools.

⁵ Collections strategy for Leeds University Library. 2011. Available at: http://library.leeds.ac.uk/downloads/file/212/collection_strategy

Many libraries are also concerned about preserving long-term access to materials for their users. This, together with the associated concern to preserve the integrity of the National Research Collection, is a prime reason behind the desire to enhance the current capability of the tools by adding retention status data. The existence of some degree of coordination at national level and the guarantee of continued access to materials within the UK makes it easier to persuade local users to support the move towards more active collection management. Indeed, the ability to use the Tools to identify “Unique and Distinctive Collections” (UDCs) can be a powerful tool for local advocacy.

Libraries also identify strategic benefits from this approach, including improved planning and the development of more coherent and systematic collections. They see how the Copac Tools can support the identification of UDCs within their 19th and 20th century holdings and the benefits (along with potential obligations) which this might bring. Equally, through confident weeding of unwanted titles, a more focussed collection which is easier to use can be presented to their user community.

However, significant barriers and risks also exist with this approach, as set out in the following table:

Barrier / risk	Response
Agreeing an acceptable definition of “long term retention”	Role for RLUK?
Lack of standards – or at any rate a lack of agreement around the pertinent standards	Propose framework as part of CCM Tools Project; support from Copac and RLUK
The need for the system to be based on reliable and quality data in order to gain acceptance	Ensure proposals are workable and imply minimum workload and risk
The extra time required to add and maintain this data	Ensure proposals carry minimum workload; revision of workflows; support from RLUK
Library reputation historically depended to a large degree on size of collections	Demonstrate how this work enhances reputation through concept of Unique and Distinctive Collections / Unique Selling Points and broader strategic concerns
Potential for bad publicity around stock disposal	Work on UDCs / USPs
Difficulty of “selling” change	Demonstrate strategic vision and value

Establishing a strategic framework for collaboration

In the last decade, there was considerable interest in formal schemes for collaborative collection management, including the transfer of unwanted material from one library to another, and CoFoR

provides an example within a single subject domain, namely Russian and East European Studies⁶. This concept was outlined at the March workshop but generated little discussion. There was some degree of support for a system which enabled a library to publicise unwanted titles and offer them to other libraries, but no concrete ideas emerged of how it might be implemented.

In contrast, one of the attractions of focussing collaboration on a looser framework that allows libraries to indicate retention status and use it to drive local decisions lies precisely in that fact – it is a framework with responsibility devolved to and under the control of the participating library. Local autonomy is broadly retained, and it can even support varying levels of involvement by participating libraries – although this could be both a strength and a weakness since it also permits any initial enthusiasm to slowly wither and die through insufficient submission of data.

The Copac database (and its CCM Tools) would provide a mechanism for such a scheme to operate, but some central focus would also be required to provide coordination and set quality standards in order for the framework to be seen as providing sufficient guarantees to make participation worthwhile. Indeed, throughout the Retention and Preservation policy workshops, stakeholders stated that such a focus is required to provide coordination at a national level to agree retention, preservation and access policies. These policies would underpin quality standards and therefore provide the degree of guarantee required for libraries to participate in the scheme, and adapt shared tools, with the necessary level of trust and confidence.

Copac, because its essential role and remit is service-focussed, would seem unable to fulfil this function. UKRR, despite playing a somewhat similar role for serials, is too much defined by its very specific remit to be able to take on the complexities of this broader role. Hence it is RLUK, as the representative body of the major UK research libraries, that is widely seen as a key player in establishing the necessary organisational framework - particularly from the perspective of inspiring confidence and ensuring quality assurance - though participation would ideally be much wider and open to all Copac contributors.

In the light of this, it is recommended that RLUK should take the lead in driving forward development of a national organisational framework capable of maintaining the integrity of the proposed National Research Collection of Monographs. This would be achieved by national agreement and/or clarification in the following areas:

- The retention status of material held by the six legal deposit libraries
- The degree to which participating libraries would be willing to be the official permanent repository for specific materials
- An acceptable definition of long-term retention for preservation as well as access

⁶ <http://www.rluk.ac.uk/node/82>

- The impact of the national backlog of uncatalogued materials in decision-making in this context
- The role of digital surrogates (digitised books; ebooks) in decision-making in this context

Such an agreement would mean that Copac could operate as a shared mechanism for a national framework to operate, for example providing functionality for participating libraries to indicate:

- intention to retain an item in the long-term
- physical condition of the item
- access conditions, including ILL status or digital availability

Although this report envisages a national framework specifically for the UK based on Copac, consideration should also be given as to whether it could be enriched or supported through collaboration with OCLC, an organisation which has been similarly active in developing tools for collection analysis and in exploring shared responsibility for long-term retention of print collections. Some form of interaction would be particularly valuable given the role of OCLC in supporting UnityUK, the public library network with a union catalogue that provides holdings information for much of that sector.

Standards and formats

A key consideration already touched on is how information about retention and preservation can be accommodated within the existing metadata context in a way which permits easy updating and supports standard use of terms. MARC tag 583 – “Action Note” – would seem to provide the necessary framework. This is already the preferred way of indicating print serial holdings retained within the UKRR scheme, so its adoption here would potentially provide a common structure for both serial and monographic holdings. 583 was also the mechanism chosen by OCLC for its “Print Archives Disclosure Pilot”, for which the final report⁷ has recently been published. This too was based on serial holdings, but attempted a tighter framework for controlling information than had been specified by the UKRR, based on fuller use of MARC 583 in association with the vocabulary defined within Preservation and Digitization Actions (PDA)⁸.

The PDA vocabulary is maintained within the MARC standards system and allows one of a number of standard “actions” to be recorded, together with the date on which the action is decided upon, further detail regarding the precise details of the action and/or condition of the item (where appropriate) and a free text note giving additional information. An important and explicitly stated consideration in

⁷ OCLC Print Archives Disclosure Pilot Project. *Final report*. 2012.

https://docs.google.com/document/pub?id=1iM86_QRG0vBXqIRwezIA2pOANJdlqmlAnSS_t31WgNU

⁸ <http://www.loc.gov/marc/bibliographic/pda.html>

recording this information is to allow other institutions to understand the condition and status of the item described in order to inform their own decision-making. Within this context, the use of a tightly controlled vocabulary permits confidence in interpreting the information recorded and would, for example, allow a library to decide not to actively conserve their copy of a title because another institution has already done so.

Originally drawn up by RLG, ARL and Library of Congress, PDA has an excellent pedigree, and it was designed to cope with use cases that very clearly overlap with our own. Because its focus is on offering the capability to record the full range of actions around preservation and digitisation, it presents a completeness and complexity which may not be helpful within a framework which is aiming to be simple, easily comprehensible and low maintenance. However this could be addressed by identifying a limited subset of subfields and terms which are capable of providing the essentials for our purposes without precluding use of the full vocabulary by any institution that so wishes, and this is essentially the way in which it was used within the OCLC Pilot.

However, PDA does not currently include a term which exactly corresponds to the concept of “intention to retain” which is so central to the current discussion (the existing PDA term “retained” is intended for a somewhat different use). OCLC recognised this difficulty and defined two additional terms (“committed to retain” and “completeness reviewed”) which it would hope to see adopted within the standard vocabulary in due course. Adoption of the same standards by both OCLC and the UK National Research Collection would clearly be beneficial to both parties, permitting data flow between the two systems and opening up the possibility of collaboration or interworking in specialist areas. Because this combination of 583 and PDA may be used in other contexts (e.g. for locally recording an institution’s decisions in the areas of preservation or digitisation, there is a risk of creating ambiguity which would undermine the confidence which is critical to successful collaboration. However this can be managed by mandating use of the \$f subfield (“authorization”, here the collaborative scheme involved). And in practice, very few UK libraries currently record any data in MARC tag 583.

It is therefore recommended that:

- work on developing the Copac Tools to support the UK NRCM should be based on MARC tag 583 and PDA;
- discussions with OCLC should take place with the objective of identifying a common approach to the implementation of these standards within this context;
- discussions with OCLC should also consider a joint approach to enhancing PDA in line with the requirements of our user communities;
- UK guidelines for the use of 583 and PDA should be established (while permitting richer capture of information by those institutions which wish to do so).

Appendix A provides some examples of how 583 might be implemented within this context.

Principal recommendations

1. RLUK should take the lead in driving forward development of a national organisational framework to coordinate and maintain the integrity of the proposed National Research Collection of Monographs. As a minimum, this would require national agreement and/or clarification in the following areas:
 - An acceptable definition of long-term retention for preservation as well as access
 - The retention status of material held by the six legal deposit libraries
 - The degree to which participating libraries would be willing to be the official permanent repository for specific materials
 - The impact of the national backlog of uncatalogued materials in decision-making in this context
 - The role of digital surrogates (digital copies and ebooks) in decision-making in this context.
2. Copac should provide the functional mechanism by which this National Framework operates and provide functionality for participating libraries to indicate, as a minimum, intention to retain an item in the long-term.
3. The framework implemented by Copac should be simple and inexpensive for libraries to operate, yet sufficiently flexible and extensible to allow additional data as required, potentially including:
 - physical condition of the item;
 - access conditions, including ILL status
 - digital availability.
4. Work should proceed on agreeing a metadata framework to support the National Research Collection of Monographs, based on MARC tag 583 and PDA.
5. The Copac Collection Management Tools should be further developed so that libraries can easily exploit this data.
6. Further consideration should be given regarding to what extent the new metadata fields could be automatically populated with data derived from that already held by Copac or within the local systems of participating libraries.
7. Discussions should be initiated with OCLC around associated areas of common interest including participation of UnityUK libraries, common metadata specifications and the scope for further collaboration around mechanisms supporting long-term retention of print collections.
8. The community as a whole should give further explicit consideration to the justification and value of retaining physical copies of items which have been digitised.

Appendix A: Examples of 583 content, based on PDA

Level 1. The OCLC Pilot permitted a 583 tag with only one subfield. For monographs, this would need to be unambiguously associated with a given institution within the shared system. This might be achieved by adding an institutional symbol in subfield \$5 at point of loading into Copac if one is not already present.

583 1 \$a Committed to retain

Minimal level of input

Level 2. A recommended set of subfields for monographs might consist of:

\$a Action e.g. Committed to retain

\$c Date e.g. 2012

\$f Authorization e.g. UKNRC (UK National Research Collection)

\$2 Source of term e.g. pda

\$5 Institution code e.g. UkleU (may be able to be supplied by Copac software on loading?)

583 1 \$a Committed to retain \$c 2012 \$f UKNRC \$2 pda \$5 UkleU

This fully complies with standard practice, albeit with a minimal level of information.

Level 3. Where a library desires to record additional information, the full range of MARC subfields and PDA terms can be used. This information might either relate to the intention to retain i.e. be expressed within the same 583; or alternatively relate to additional information about the item held i.e. be expressed in one or more subsequent 583 tags. These 2 possibilities are shown in the two following examples.

583 1 \$3 v.1 only \$a Committed to retain \$c 2012 \$d 2022 \$f UKNRC \$f SCURL \$5 UkleU

In this case, the Library has committed to retain vol. 1 only, for a 10 year period, within 2 separate agreements: UKNRC and SCURL

583 1 \$a Committed to retain \$c 2012 \$f UKNRC \$5 UkleU

583 1 \$a Condition reviewed \$c 2012 \$l mutilated \$z pages 9-15 damaged \$5 UkleU

583 1 \$a Housed \$c 2012 \$i box \$5 UkleU

In this case, the Library has committed to retain the item. On reviewing the condition, it has noted that the item is mutilated and has also decided that it needs a storage enclosure (to be housed in a box).