Collections and the Archive Layer

In this brief paper I want to concentrate on the major component of the collections of research libraries – the books and journals, printed and electronic, that constitute their scholarly research collections. I am not therefore talking about special collections, which are essentially of institutional value, and inevitably partial, rather than an attempt at completeness which is what our general collections traditionally aspire to be. I have also been tempted by the topic to consider the changes in library responses to the challenge of collections across the timespan of my own career – from the mid-1980s onwards - since that timespan has seen the arrival of the internet and the grappling with its consequences which have required libraries to consider questions of their values and principles. This has of course not been a bad thing, but we have not yet reached any shared agreement on the question of core library values and principles, with so much of our practice having been shaken up by the internet and the way it has changed publication and communication behaviours. And we do not yet have an agreed and uniform mode of practice in the research library sector in respect to collections, nor even a way forward towards one.

Research libraries are now a lot less likely to behave like each other than they did back in the 1980s and before. Most of us are sending journals to UKRR – but some at much greater volumes than others. Why is this? And what about our scarcely-used books? Most of us have library stores, but what we put there as opposed to what we discard will vary dramatically. Some of us have dedicated library spaces for researchers, and feel it imperative that they contain books; others have similar facilities and feel it imperative that they don't contain books. We all have repositories, but there is no benchmark measure that indicates what is a good repository for a research library.

In the pre-web world, the idea of comprehensiveness within the walls of individual research libraries still existed. Research libraries then collected the world's scholarly literature at a time when it was considerably easier to do so than it is now. Most of that literature was published, either as monographs or journals, and arrived in libraries in the form of book-like objects, where, once processed, it could take its place within miles and miles of shelving. Collection policies, if they existed, could be fairly straightforward. National and deposit libraries collected everything published within the country, and, in the case of the largest libraries, a lot more besides – eg the major scholarly works of different world literatures. The friendly web clip on the home page of the Library of Congress website says 'if it's culturally significant, and useful for research, chances are we have it.'

There was, we acknowledged back then, the 'problem' of grey literature, but it was an esoteric problem. Libraries with a particular interest, say, in the reports of an international agency that lay 'outwith bibliographic control', as we also said then, might elect to collect in some of these areas, but this was activity which was very much marginal. It was a form of special collections, without the preciousness. Today we inhabit a world in which bibliogaphic tools have lost the authority that once allowed fugitive material to be classed as 'grey', and unpublished material is often as valuable as published, of which it is sometimes an alternative and approved version. Or, to put it another way, the 'problem' of grey literature is now endemic.

Within the walls of our research libraries we held a representation of the complete scholarly archive. It was a version of the archive which was essentially good enough for our researchers. And what we didn't hold, we knew about because we had records for it. But what we have seen, in our career lifetimes, is a loss of that institutional-level control and completeness to which back then we still aspired. Librarians were controllers, and libraries were obviously controlled zones. Look at the way Ross Atkinson of Cornell University Library, one of the major Collections Librarians in American librarianship, talked about this back in the earliest days of the web, in 1996:

> The network is not a digital library. We cannot sit back and imagine that what is on the network is in the digital library … A library, digital or otherwise, is always a highly selective subset of available information objects, segregated and favored, to which access is enhanced and to which the attention of client-users is drawn in opposition to objects excluded … it is time - past time - for the academic library community to begin work on the creation and management of a single, virtual, distributed, international digital library, a library that has (conceptual, virtual) boundaries, that defines its service operationally on the basis of the opposition between what is inside and outside those boundaries, and that bases that service on the traditional social ethic

that has motivated all library operations in modern times. The academic community must consider, in other words, the creation of a control zone …when an object of information is moved across the boundary from the open zone into the control zone, then that should be done with the understanding that the library community takes certain responsibilities - and makes certain guarantees - for the quality and accessibility of that object indefinitely…That epistemologically and ethically essential function of the library in the world of primarily paper information must be retained and strengthened as society moves increasingly into the online information environment.[1]

Atkinson died 10 years later, shortly after running a major conference at Cornell on the future of research collections (the Janus Conference). By then, though he no longer called it the 'control zone', he was still talking – quite presciently in some ways - about the same idea:

> Within this category we locate our fifth challenge for collection development—the enormous one of archiving. This challenge must be approached in two parts, print and digital. We will require decades, generations, to move our paper holdings into digital form. In the meantime, the maintenance of large warehouses of print materials will become ever more costly. It is essential, therefore, that research libraries divide among themselves responsibilities for archiving low-use print materials. With respect to digital information, the most serious challenge universities and their research libraries face is how to reappropriate the responsibility for the preservation of key scholarly objects that are now maintained primarily or exclusively on the servers of publishers and other vendors throughout the world. Technical, economic and even political impediments can jeopardize continued access to such objects, despite the best intentions and commitments of publishers and vendors. It is essential therefore that research libraries re-assume full responsibility for archiving such scholarly materials for the long term.[2]

Are we still controllers? Are our libraries controlled zones? It may not feel like it, and we are more likely these days to wince at the idea than we would have even 10-15 years ago, but surely we still are? And don't our researcher users still expect us to be? Only now we have to do it without the concept of 'universal bibliographic control', a major IFLA objective for many years, which has been surrendered in the digital deluge. Hence the need for our collections policies of recent years, laying out what we do and do not collect as institutional libraries, in a way that justifies a retreat from completeness.

Yet our policies reassure our researcher colleagues, that, though institutionally we are necessarily compromised in all sorts of ways – by lack of space, by the way publishers sell and license content to us, by its impossible cost rises, and just by the general sense that collecting is difficult and fragmented in our day - comprehensiveness is still available. It's just that it is provided in a complementary way by the 'system' as a whole, into which we are plugged. Out there – as has been true for decades - is still the BL, and our research library partners which will supply on ILL. But increasingly the layer that supports our own printed and licensed collections is being built up from other sources too. This is being achieved through the mainly uncoordinated efforts of several different organisations, and is composed of content in both print and electronic formats - UKRR, Portico, CLOCKSS, Hathi, the stored print collections of various US regional cooperatives, and major research repositories. On top of this archive layer is a layer of metadata, generated largely by libraries themselves, but also by publishers and other agencies, and husbanded by OCLC and others.

Some of these collaborative efforts represent the fruits of library thinking about preservation, and some the fruits of library thinking about the management of redundancy. And indeed, there is a relationship between redundancy and preservation which is changing as we address this question of the archive. Lorcan Dempsey recently put the *status quo ante* situation typically concisely: 'Preservation is a benign artifact of the print

---

[1] Atkinson, Ross (1996) 'Library Functions, Scholarly Communication, and the Foundation of the Digital Library: Laying Claim to the Control Zone', *Library Quarterly*, v66 n3 p239-65.
[2] Atkinson, Ross (2005) 'Six Key Challenges for the Future of Collection Development' (Janus Conference) in e-Commons@Cornell

publishing model as materials are redundantly available across the library system.'[3] Except that whereas then the redundancy was 'passive redundancy' in the environment of 'benign neglect' to which Dempsey alludes, with duplicated printed materials insecurely held on open shelves in a large number of research libraries, now it is increasingly 'active redundancy', with a smaller number of duplicated items more securely held in stores. And that environment lends itself to collaborative management, as we see with UKRR and the various US collaborative print depositories.

While I worked for OCLC Research I was quite involved with a programme it ran on Research Information Management. One of its major objectives was to understand the world of researchers, from the centre out, with the researcher at the centre. We reached a point at which we could see four of these environments quite clearly: the environment of the researcher's own domain; the environment of their institution; the assessment environment (represented in the UK mainly by REF); and the funding environment. What was not nearly so clear, though we wanted somehow to see it, was the environment in which the materials of their domain – those they consulted and those they produced - were maintained and preserved. In other words, the scholarly archive. Back in pre-digital days, that would have been the contents of research libraries – what they held on their shelves. The content of the archive – represented materially by the physical collections - and the means of access to it, were largely identical. Now the archive is not constituted by what is held within the Library's walls, nor even by those holdings plus the licensed content it provides from the cloud. These collections are no longer thought to constitute a complete, guaranteed and permanent store of scholarly materials proofed against loss. They are not an underpinning layer of the academy. Our institutional libraries are in retreat from their role in providing the scholarly archive.

However, they can of course contribute to the archive layer. Part of the layer takes the shape of UKRR. Part of it is in the shape of Portico and CLOCKSS. Part of it is WorldCat. Part of it is UK PubMedCentral. Part of it is Hathi. Lorcan Dempsey, again, writing recently said: 'Think of Hathi Trust. A few years ago, it is likely that libraries would individually build infrastructure to manage digitized books and store them locally. It is now accepted that this is better handled in a consolidated way, gaining from economies of scale, but also from being able to put a unified resource on the network.'[4] There will be other parts.

But we don't yet see this emerging infrastructure as an archive layer, and we are as far from having any governance of it as we were in Ross Atkinson's day. I think we need, as a research library community, to take one step back from our institutional libraries, and then one further step back from our national perspective, and think more about how to create it and direct it collectively and efficiently. It is of course a diverse patchwork of services. It consists of services provided by cooperatives, national agencies, national libraries, publishers, disciplinary hub services and content archive stores. Part of it is operated by private, but non-profit organisations, all of which claim to operate on behalf of libraries and research. Can the international research library community find a way to influence and shape it? Can it consider where it is collectively deficient, where it has pieces missing, and how those might be remedied?

This archive layer, or environment, branded with an international research libraries imprimatur, ought to be strong enough to influence assessment requirements, research funders and disciplinary domains themselves. The domain requirement for good archival behaviour should directly impact the researcher, just as the mandates from the bodies handing out grant monies and funding (and ranking) institutions via assessment exercises will impact institutions. In this model, the research library influences the development of the scholarly archive. This archive layer needs both to provide a permanent preservation repository and to have ways of articulating with local, institutional library services. It must hold the content that comes to the library 'from the outside in' as well as the stuff that is generated 'from the inside out'. It is what I believe Ross Atkinson was calling for as he saw his collections impacted by publishing on the web. It is the developing future of research library collections.

John MacColl, University of St Andrews. October 2012.

[3] Dempsey, Lorcan (2012) 'Libraries and the Informational Future: Some Notes' in *Information Professionals 2050: Educational Possibilities and Pathways*, ed. Gary Marchionini and Barbara B. Moran, 113-26. Chapel Hill, NC: University of North Carolina at Chapel Hill.
[4] Ibid.