Refactoring and reframing collecting in a digital age
David Pearson

The first thing I want to say is that I think this is an important topic, and that thinking about where we should be going in developing special collections in our increasingly digital professional landscape really does matter. You might think that's stating the bleeding obvious, but there has been a lot of shifting of emphasis in recent years in statements of professional intent or purpose for libraries and repositories and I think it's not by any means any more enshrined in professional belief that collections are important. With more and more material becoming available electronically, and access models for libraries relying as much on licensing as on owning, librarians have increasingly been reinventing themselves by focusing more on services, repository management, the skills they have in navigating their way around information, and in training users in information literacy. Big used to be beautiful but that can no longer be taken as read.

When the Consortium of University and Research Libraries was set up in the UK in the 1980s its membership was obvious – it was the half dozen or so university libraries with the biggest collections. Anyone who sat through the numerous discussions which took place at the Board of CURL, or Research Libraries UK as it now is, during recent years – and there are several of us in this room – will know that such factors are now much less valid. The idea of size alone as a criterion for joining such a group is now considered an irrelevance bordering on the politically incorrect. Is the Bodleian Library a better research library than Leicester University Library, because it has more books in it? They both offer access to a huge range of online content and if you're looking for the text of a rare 16th century English book, they both provide access to *Early English Books Online*. Another facet of this glass can be seen in the switch of emphasis in many curatorial jobs – in the museum sector as well as the library and archives one – from traditional collection-building to the exploitation and interpretation of what's already there. I think this is one aspect of the changes that have been taking place at the British Library this year, though as I say museums have been redefining their activities in this direction for some time.

I do think that exploitation, interpretation and the use of digital technology to open up access are hugely important, and I think that roles and responsibilities do need to be regularly appraised in this regard, but I will nail my colours to the mast and say that I still believe that underneath all this, you need the stuff to be there. Collections matter, and I think they matter to users rather more than librarians sometimes think they ought to. The library at Alexandria was celebrated across the ancient world not because it had such friendly librarians or a really good café, though I'm sure it had all those things, but because it held so much of the world's recorded knowledge. For centuries, libraries have been sustained around that logic – their fundamental purpose has been as places where knowledge and information is stored and made available, and that has been what has drawn users to them. I no longer run an academic library but rather a portfolio of services which does everything from lending Mills and Boon to providing medieval letter books for PhD students, but whatever people are after I sense they are still drawn primarily by the expectation of finding content, whether it's a novel or a textbook or a medieval manuscript. While they value those other services, they would be less enthusiastic if the content was taken away. The British Library's strapline, the world's knowledge, continues to trade, quite rightly, on the fact that its very large collections embrace a significant proportion of that knowledge. Of course, the stuff can be digital instead of written or printed; the key thing is that the users can get at it. I noted with interest that the British Library's new 2020 vision, published on its website last month, includes the statement that they will move to “collecting digital content as the format of choice, actively making the print to digital transition”.

So what is the future for special collections in this evolving digital world? The point is often made, and I think rightly so, that it’s a more enduring one than seems to be round the corner for more general collections, which are likely to look less sustainable in their traditional forms as mass digitisation programmes progress. It’s an argument advanced by Robert Darnton in his recent *Case for Books*, so I know I’m in good company. The UK Research Reserve is showing the way to the pulping mill down which many miles of printed paper are likely to travel once digital technology for reading texts is thoroughly bedded in, though most people think that will still take a while to happen. Let me remind you, though, that Amazon are now selling more e-books than hardbacks. Whatever people choose to do regarding their personal ownership of books, the arguments for investing large recurring sums in maintaining paper mountains at public expense will become harder to defend. The case for the prosecution will of course rely significantly on the premise that multiple copies of printed books or journals are identical, which was the whole point of Gutenberg’s invention, and for modern printed materials this is generally true. Two copies or one hundred copies of the latest report from OCLC Research are all the same when they come from the printer, the book format is a convenient device to enable the storage and communication of the content, and if that content is equally easily available online it’s hard to see the need to keep lots of paper copies. The stuff is still there, and accessible, as far as the users are concerned, and that’s what matters to them.

Special collections are different; they contain material which is rare or unique, because it’s old, or written, or handcrafted, they contain material whose research interest extends beyond just words or pictures into the artefactual nature of the media, they contain things which are recognised as belonging together because they were brought together in that way. They can still be made available as digital surrogates, and continuing to open up access to special collections through digitisation is a hugely important agenda, with complex questions. That’s beyond the remit of this paper but features elsewhere in this symposium’s discussions. Even if digitised, it will remain absolutely necessary to preserve the physical originals of unique material, because the digital version remains a surrogate and we may wish or need to return to the original for various reasons. It follows that the special collections will increasingly become the features which distinguish research libraries; if Google’s march through the printed output of the 20th century continues, and the copyright access issues are suitably steamrollered, as will surely be the case, the things that will distinguish the Bodleian from Leicester will be less the copyright receipt collections and more the John Johnson Collection, or the medieval manuscripts, or the literary manuscripts of Alan Bennett. The holding of such material will not only provide quarries for researchers, it will also enhance community recognition and provide opportunities for revenue generation through digital and other means of exploitation.

What, though, is uniqueness, even for special collections, in this digital landscape? It is, most obviously, material which literally and obviously exists as only one exemplar – manuscripts, archives, drawings, some of the more fugitive kinds of printed material like ephemera – the John Johnson Collection here in Oxford is a good example. Staying in Oxford, what about, let us say, the books of John Selden, the 17th-century lawyer whose 8000 books comprised one of the largest private libraries in England at the time of his death in 1654, and which was one of the major early acquisitions of the Bodleian Library? Selden End, in Duke Humfrey’s Library, perpetuates his memory. Selden’s books are unique in that they were his, they may contain his inscription and marginalia, they may contain evidence of other ownership or annotation, they are all individually bound. As texts, they are less unique; the point about Gutenberg's invention applies to them too and there are very few books within that 8000 or so which are unique surviving copies of their particular editions. They have for many centuries been valued as a research resource because the textual content they offer has been hard to find elsewhere – they are a concentrated pool of genuinely rare books – but the kind of access which digitisation is opening up is bound to weaken that logic. Here is Thomas Gataker's work on dipthongs, a mid 17th c English book which is known to exist, according to ESTC, in 7 copies around the world, 6 in this country and one in America. One of those is Selden's in the Bodleian and it was indeed the copy filmed for EEBO. Until recently, 17th-c philologists would be likely to travel to Oxford to see this but now they are far more likely to consult it electronically, if the text is what they are after. Selden's copy will generate a little sliver of EEBO royalties for the Bodleian but the book may be called from the stacks less often. If one of the other 6 libraries decided to digitise their copy in order to provide a textual surrogate they might reasonably be said to be reinventing a wheel.

The key point here is that the textual function of printed books of all periods is likely to be increasingly taken over by electronic means, which will be regarded by many users as satisfactory. Indeed, not just satisfactory, they will be downright delighted. EEBO was marketed a few years ago with the slogan "it's like waking up in the stacks of the British Library" and the convenience which it has brought to many researchers' lives is transformational. I have recently been doing some work on late 17th c auction catalogues and the usefulness of being able to call these up onscreen and print them off, without travelling around libraries, hardly needs describing. For what I was doing, the surrogates were 100% satisfactory. Arguments about the importance of continuing to use physical originals are sometimes just nebulously sentimental, but sometimes based on the rather more concrete logic of needing to experience the 3-dimensional reality of books in the way their original users did. We shouldn't overlook the questions around the impact of physical form on communication, and the relationship between medium and message, but nor should we deceive ourselves that the conditions in which most people experience early printed books are anywhere near what their early owners knew. We tend to think of old books as things that are fragile, precious, in need of careful treatment and only accessible in special reading rooms with bookrests and snakes. If we want to recreate the environment of their first owners, we should get to recognise what leather was like before it was brittle and cracked, and imagine handling them less reverentially. To put it another way, the argument that you need to read *Paradise Lost* in a late 17th century edition in order to recreate the experience of a contemporary reader needs more than producing a fragile copy to be read with snakes and supportive foam in order to be genuinely authentic.

But let us not forget that there is more to books than words, and that their properties of specialness and uniqueness, and their research value, may lie elsewhere. Guildhall Library has two copies of the Latin edition of the Nuremberg Chronicle, printed in 1493, an illustrated history of the world which is one of the best-known incunables and arguably the world's first coffee-table book. One is a fairly unexceptional copy, with very little added to it after it left the printing house, in a late 19th-century binding. The other has contemporary hand-colouring throughout, extensive marginalia, and was owned not long after it was printed by a London citizen who added a long list of Lord Mayors and sheriffs from 1189 to 1508. It has a note to say it was bought in 1495 and we have just discovered that the original owner was Robert Fabyan, well known as a compiler of chronicles and an early London historian. It's interesting in all kinds of ways as a manifestation of early 16th-century London history and of contemporary attitudes to, and access to books. Its value lies not so much in what is put on the pages in printer's ink – for which a digital facsimile of the copy in Munich is now freely available on the web – but for what was put there in many other kinds of ink.

Our libraries are full of things like this, whose research value is increasingly being recognised by people who are quarrying the evidence to explore the social impact of books in all kinds of ways. The history of the use and reading of books, the value of marginalia, the lessons to be learned from looking at collections as collections, are all areas where scholarly activity has been burgeoning during the last decade or so. I've written more about this, about the many ways in which books may have research value as objects beyond their textual content, in my 2008 book *Books as History*.

What does all this mean for our collecting policies? As I said before, I think that special collections continue to be a unique and important strength of libraries and the institutions they belong to, for reasons beyond the purely altruistic ones of advancing human knowledge. The expansion of interest away from the purely textual to the artefactual and book-historical has had a considerable impact on cataloguing practice during recent years, and we now see libraries including copy-specific information in their metadata as a matter of course, something which was much less the case 20 years ago. Have we seen a similar evolution in collecting patterns? Less so, I think. The traditional approach to mainstream collection development activity for special collections in academic and research libraries is typically primarily textual, with a focus on filling gaps. Copyright libraries will acquire early imprints which slipped through the cracks in centuries gone by, while other libraries will go for material which has a strong subject fit with particular areas of strength.

I tested this out by looking at a few special collections development policies as found on the web, and found it quite easy to support that hunch. The British Library's stated policy for British Collections 1501-1914 is to acquire, in this order of priority, authors new to the catalogue, works new to the catalogue, substantive editions and finally other editions of works already held. The University of Manchester defines a series of subject areas as categories of special importance, including theology, the history of science and industrial history, and explicitly states that duplicate or variant editions are not normally acquired solely on account of provenance, although this may be considered in exceptional circumstances. Birmingham University includes "rarity of copy" as one of its criteria although it does recognise the potential value of physical characteristics in order to support the academic study of the history of printing and publishing. King's College London focuses its policy around a list of subject areas, and states that duplicate copies of editions already held will not normally be acquired, unless something has a significant provenance.

Those are just a few examples drawn from what's readily available on the web, and I'm not out to criticise any of those institutions – perhaps the reverse, to praise them for having such publicly available policy statements on their websites. There are numerous other Collection Development Policy statements for special collections for libraries from both sides of the Atlantic to be found on the web, which could be used as further examples. These statements often include a reference to working with other libraries, and taking account of other copies which may be locally available. If a rare 18th-century English pamphlet comes on the market which is held in, let us say, Cambridge, Durham, Edinburgh, New Haven and Chicago, but not Oxford, but which is available in full-text through ECCO, should Oxford buy it? I can see arguments around the completeness of the printed archive and around the value that may be derived from having a physical copy here, but suppose it's a choice between the pamphlet and a copy of a relatively common 18th-century book which has been extensively annotated by a contemporary undergraduate? Which has the higher quotient of uniqueness and research potential? Is "we've got a copy of that book already" a relevant consideration? If we've got copies of *all* these books available already, in online versions, doesn't the text-focused collecting policy become more fundamentally flawed?

I think that some of our smaller and less geographically central historic collections are a bit ahead of the game here, in recognising the shifting emphasis behind the value of their holdings. National Trust libraries, for example, are used to the fact that no-one is going to travel into the depths of the countryside to read a book that's more easily available elsewhere, but are increasingly understanding that the things which bring researchers are the individual histories of their copies and collections. I also know of examples of items or collections being taken in deliberately for their copy-specific rather than their textual content, like the Rosenthal Collection of heavily annotated renaissance books acquired by Yale a decade or so ago. If you read through booksellers' catalogues, though, you will get the impression that there is a still a strong emphasis on the rarity of editions as a criterion to tempt purchasers – "only 6 copies in ESTC", "not in Wellcome", etc. I think the mindset of the antiquarian book trade, who comprise an important partnership arm for the development of special collections in libraries, is still largely rooted in the collecting for texts – if you look at the way the search interface on abebooks is structured, that becomes clear.

Am I oversimplifying or overstating the case? There is an irrefutable logic to the thought process that goes: we acquire books primarily to make their texts accessible; texts have become readily available online; we no longer need to acquire books. Do I think we should stop acquiring books for special collections according to that traditional logic? Not overnight, but I think we should give careful thought to the motives behind our acquisition policies and to what kind of material will genuinely enhance our collections. We should tune our antennae carefully to emerging directions of book historical research and we should be able to take some leads in helping to develop new directions of academic thought. We should think carefully about our cataloguing of special collections, and where the effort goes in creating metadata to make this material discoverable. We should avoid reinventing wheels with traditional detailed transcription of titlepages which are now readily visible in digital facsimile, and concentrate more on the things which make these books genuinely unique. We should continue to understand the importance of special collections for underpinning research of all kinds, but we should recognise that their uniqueness and value may come to be found in different areas from those traditionally associated with books, and adapt our collecting policies accordingly.