

Libraries, the Long Tail and the Future of Legacy Print Collections

Paul Genoni
Curtin University of Technology

Introduction: the long tail

Since Chris Anderson first aired his concept of the emerging “long tail” market in an editorial in *Wired Magazine* in 2004, librarians have been contemplating its relevance to the future of information storage and access. Anderson’s long tail theory is an attempt to explain the changes that have occurred to markets as both the records of trade (that is, inventories and catalogues) and the traded items themselves, have become digital rather than physical (or in Anderson’s terms, bits rather than atoms).

In particular, Anderson stresses the emergence of the economic value to be found in what he calls the long tail. The long tail is constituted by those items which would not find a commercially justifiable place in the physical stock (the “head”) of a high-street retailer, but which nonetheless form an important part of the low-cost storage and distribution economy that is made possible by the Internet. The greater the use of digital technology for promoting, storing and distributing goods the lower the cost structures will be, and the more products will emerge in the long tail. Anderson argues that this new economy is being driven by consumers who are able to buy from the “niche” items made available in the extended marketplace of the long tail. Accordingly the future of retail will be marked by abundance, and although the head will continue to account for the bulk of sales, consumers will also be able to move ever further down the “unlimited selection” available in the long tail.

Unlimited selection is revealing truths about what consumers want and how they want to get it in service after service, from DVDs at Netflix to music videos on Yahoo! Launch to songs in the iTunes Music Store and Rhapsody. People are going deep into the catalog, down the long list of available titles, far past what’s available at Blockbuster Video, Tower Records, and Barnes & Noble. And the more they find, the more they like.
(2004)

Anderson’s *Wired Magazine* article did not consider the case of libraries at all, and libraries received only a cursory mention when his thoughts were expanded to book length in 2006 as *The long tail: How endless choice is creating unlimited demand*. Library bloggers nonetheless picked up on the long tail concept and began discussing how it might be applied to their own domain of collecting, storage and distribution. For the most part the reception was favourable. Many bloggers were keen to point out that by collecting and storing the uncommercial, the little known and the out-of-print; and by meeting demand via inter-library lending, libraries have long been in the long tail business of servicing niche markets. The idea of the long tail soon entered the mainstream of library thinking, with articles by Tom Storey (2005) in *OCLC Newsletter*, and Lorcan Dempsey (2006) in *D-Lib Magazine*, and Anderson appeared as a featured speaker at the ALA Annual Conference in June 2006.

Storey and Dempsey both linked the long tail concept to the issue of the future management of print collections. Storey's article incorporates interviews with several research library managers, perhaps most tellingly with Robert H. McDonald of Florida State University. McDonald points out the increasing tension between digital and print collections when it comes to providing service to users.

"I am curious about the effect that The Long Tail will have on our book collections . . . Most users prefer online resources because they can be accessed anywhere. But the maintenance and upkeep of legacy book stacks in many ways prevents the further extension of what libraries are trying to do with electronic resources." (Storey, 2005)

Dempsey commences his paper by pointing out "the real relevance of the long tail argument, which is about how well supply and demand are matched in a network environment", but he nonetheless focuses on the aggregation of demand through the building of federated finding aids and leaves open the question of exactly how supply might be aggregated. Notably, however, Dempsey advocates the need for "an agreed policy framework" with regard to developing storage options, and discusses how services might increasingly be scaled-up to system-wide operations. He concludes that, in order to achieve "system-wide efficiencies" what are needed are "new services that operate at the network level, above the level of individual libraries" (Dempsey, 2006).

To date these discussions of the long tail have largely accepted without question the veracity of Anderson's model and its applicability to library collections. The purpose of this article, however, is to examine some elements of the model in order to firstly, determine if it can indeed be simply adopted for use when describing the library collecting environment; and secondly, if there is some divergence, what does this tell us about the future management of print collections?

Repositories and legacy print collections

Before returning to that discussion, however, it needs to be acknowledged that now is an appropriate time for librarians to engage with the question of the long tail. At the same time that the long tail was beginning to impact on markets and Anderson was formalising the concept, research libraries have been grappling with the problem of how to manage the space consuming but rarely used items (the library long tail) that not only dominate their shelves but are also becoming increasingly marginalised as information seekers are attracted to digital content.

There has no doubt been relief in many libraries—research libraries in particular—that the increasing acquisition of digital content has promised some respite from the constant demand for new storage space. To some extent this promise has been realised, but for other reasons associated with the expanding digital content the pressures on space have hardly abated. For example the demand for more publicly accessible terminals is often met by encroaching on spaces that once housed print materials, and libraries have taken to implementing space-intensive "social" areas where users can gather for a shared networked learning experience.

To date the advent of digital content has therefore done little if anything to reduce the space pressure on research libraries—a pressure which was created initially by their concerted effort to provide local access to the long tail of research material. At the same time there have been signs of a growing concern for the future of the long tail as it has become apparent that for some libraries their response to the pressure to deliver collections electronically may result in print materials being dealt with expeditiously and without proper attention being given to established standards of conservation, preservation or retention.

These circumstances have led to the emergence—or more accurately the foregrounding—of the concept of the shared repository as the vehicle for storing, managing and making available the long tail that is constituted by legacy print collections. The idea of shared storage facilities is far from new, but the increasing expense of local retention; the enhanced possibilities for rapid (including digital) transfer of stored items; and the emerging demand gap between digital and print material, have converged in such a way that the arguments for repository storage have become increasingly compelling.

The repository solution is grounded in a dual vision of the future of research libraries. Firstly, that the interests of individual libraries may be best served by reducing their responsibility for retaining and managing little used material; and secondly, that a library “system” as a whole—be it regional or national—may be advantaged by the amalgamation of residual collections of print material in order to achieve economies of scale with regard to storage and supply.

National repositories of research materials had been established as early as the late 1980’s in Finland (Vattulainen, 2004; Saarti, 2005) and Norway (Henden, 2005); and the 1990s in France (Vattulainen, 2004). Other countries, such as Australia (O’Connor, 2004) and Germany (Kempf, 2005), also made tentative and less systematic gestures towards the development of a repository solution by implementing forms of regional storage.

In the United States there is as yet no coordinated approach to the development of shared regional stores, but there is evidence of increasing interest in the concept. This interest was made apparent when the Task Force on the Artifact in Library Collections, operating under the auspices of the Council on Library and Information Resources (CLIR), issued its final report recommending the development of a network of “regional repositories”. The Task Force report called upon librarians and researchers to act collaboratively and:

Advocate for the development of regional repositories of artifactual collections that reduce duplication of effort, create economies of scale, and ensure that the greatest number of unique or scarce priority items are preserved and made accessible to researchers (Task Force, 2001).

Building on this report, CLIR commissioned a second report which provided a far more detailed analysis of the issues involved in developing and implementing such shared repositories. This report concluded that:

With the appropriate resources in place, one could imagine the major North American research libraries, regional repositories, and national-level repositories linked in a network that enables strategic management of the important primary resources for scholarship. (Reilly, 2003, p. 40)

At the same time that the CLIR reports have promoted the development of repository storage, practical evidence of growing US support for the concept has been provided by the development of a number of shared storage facilities by collaborating university libraries recognising the advantages of such amalgamation (Payne, 2004).

Activity in the United Kingdom has been more intense and systematic. Scotland implemented a national academic library store—the Collaborative Academic Store for Scotland (CASS)—in 2004 (Nicholson, 2004). CASS was developed as a joint initiative of the Scottish Confederation of University and Research Libraries and the National Library of Scotland. This joint activity between academic libraries and a national library has been mirrored in England, where the lead in repository development has been taken initially by the Higher Education/ British Library Task Force reporting to the Research Support Libraries Program, and subsequently the Consortium of Research Libraries working in conjunction with the British Library and the Research Information Network.

The key outcome of the UK activity to date has been the completion of the report *Optimising Storage and Access in UK Research Libraries* (Chems Consulting, 2005). This report offered a number of ways forward for the development of a national repository for research materials, and following further consultation and national meetings (Evolution or revolution, 2005) a consensus has emerged for the so called “Option 4”, which is based around a repository system—the National Research Reserve—built on the collection of the British Library.

It should be noted, however, that these developments favouring shared repositories as the model for research library storage have largely taken place without the benefit of, or reference to, Anderson’s influential analysis of the long tail phenomenon. At first glance, however, it would appear that repository based solutions to library storage problems are using some of the same logic that influences commercial retailers of long tail markets. That is, by aggregating demand through massed digital inventories (national and other forms of union catalogues) where every item has an equal claim on the attention of the user, and aggregating supply in a manner which allows for the most rapid and cost-effective distribution, they are enabling access to, and reducing the cost of, their particular long tail.

I will return to the issue of repositories and their future role in the management of print collections, but it is first necessary to examine in a little more detail exactly how well (or otherwise) Anderson’s version of the long tail describes the circumstances of research library collections. Indeed, what I wish to highlight are several ways in which a research library collection can be distinguished from the Andersonian long tail.

The long tail and popular culture

Firstly, it is apparent that Anderson's long tail model functions in its purest form with items designed for mass consumption, and in particular popular culture material. This may seem at first a little counter-intuitive, as the long tail caters for niche rather than mass interests, but these niches can only emerge in markets that are supported by very large numbers of consumers. It is telling that almost all of the examples Anderson draws upon are from popular music and film, and indeed his original *Wired Magazine* article ran under a banner proclaiming the long tail niche markets as "the future of entertainment" (Anderson, 2004). Of course most pop culture production is not overtly popular—to be so would place an item in the head rather than the tail—but virtually all of it is at least potentially of interest to fans of particular genres.

The concept of genre is emphasised repeatedly by Anderson and is a critical difference in explaining why popular culture and research material produce different long tail effects. Genre-based fandom is a personal interest pursuit, and fans become in effect "collectors" who are not *primarily* concerned with issues of quality or adherence to any externally imposed standards of production or market appeal. It is in the nature of many pop culture consumers to actively seek out niche items as a means of identity creation and as a conscious search for something different from the content found in the head. Genre-fandom creates a marketplace that values novelty and innovation, and in which individual consumers often create their reputation amongst fellow genre aficionados by discovering the most obtuse and marginal material to be found in the long tail.

It should also be noted that Anderson grounds his model in the pop culture notion of "hits" and "misses", with the former going on to constitute the head, and the latter the tail. He points out that in such a market the creation of hits is often the result of media access and marketing power and it is a process that is frequently remote from objective measures of quality.

For too long we've been suffering the tyranny of lowest-common denominator fare, subjected to brain-dead summer blockbusters and manufactured pop. Why? Economics. Many of our assumptions about popular taste are actually artifacts of poor supply-and-demand matching—a market response to inefficient distribution. (2004)

According to Anderson, notwithstanding its apparent value, a lot of the "quality" material is doomed to end up in the long tail.

This is quite a different process to that which applies in the use of research collections. The mechanisms of quality control that are universally practiced for academic publishing—peer review; editorial independence; journal ranking and reputation; library selection—are intended to ensure the identification and selection of quality material. The outcomes of this process may not be as predictable as one might hope and it undoubtedly fails from time-to-time, but it almost certainly has the effect of producing a greater concentration of quality in the head than is the case in pop culture markets. And as successful research is frequently built upon previously successful research, the individual researcher has little or no incentive to spend time extracting material from the tail.

In other words the concept of genre as discussed by Anderson is not analogous to the notions of “discipline” or “subject” which form the basis of academic research. Academic reputation is far more likely to be built on a comprehensive and expert knowledge of the material found in the head than in a familiarity with the research long tail.

A related distinction between popular culture items and others—in particular research resources—is to be found in the various systems of recommendations or “filters” that Anderson describes as being crucial to locating items in the long tail. He makes the point that filters such as genre and sub-genre categorisation; consumer pattern matching; customised playlists; editorial recommendation and consumer reviewing, are the critical means by which pop culture fans are led through the mass of content in the long tail. As a result the long tail of popular culture has become a “space” where consumers go to search for content as a leisure activity.

With academic material, the search process works quite differently. The principle method of retrieval—be it from the head or the tail—is the “known item”. When the long tail is searched more generally it is rarely a process that can be assisted by the styles of fan-based filtering that can function effectively with popular culture material. Perhaps the nearest equivalent is the citation-influenced ranking employed by the likes of Google Scholar, but even this is a comparatively crude filter when it comes to reflecting the real matter of interest to users of research material—the intellectual content.

The long tail and the commercial marketplace

Secondly, the long tail view of retailing as described by Anderson applies in a commercial environment, where digital retailers (“e-tailers”) are driven by desire for profit and consumers are motivated by their desire to acquire niche items. The long tail model is essentially an economic theory grounded in fairly traditional market forces, albeit with an overlay of new technologies. It is apparent, however, that research libraries have long existed outside the commercial market, and even if they are becoming subject to greater accountability pressures, their performance measures are not those associated with profit-driven businesses. Indeed it is because libraries were outside the market economy that they were able to develop their role as collectors and suppliers of the long tail at a time when retailers could not.

Anderson describes the “big economic trigger” that started the long tail wagging as being the reduction in “the costs of reaching niches” (2006, p. 53). These costs, which have previously been passed on to consumers in the shape of higher retail prices, have been reduced by leveraging the economies of massed digital demand and supply. For many consumers of library sourced information, however, the costs have been largely transparent in that the transaction is not a commercial one, and even if the library user has been required to make a contribution (as in the case of an inter library loan) it has rarely reflected the full cost of the service received.

Libraries have done a significant amount to reduce the opportunity costs (time and effort) associated with long tail discovery and retrieval—and according to Dempsey there is still more to be done—but for most users the financial barrier has not been the issue that it has been in commercial markets. By an extension of Anderson’s “trigger”

logic, there is therefore less scope for growth in demand bought on by falling prices for library long tail material than is the case in the commercial sector.

The long tail and the persistently physical

Thirdly, Anderson stresses the importance of the digitisation of a product for distribution if the long tail possibilities of a market are to be realised. As he explains:

Digital catalogs of physical goods lower the economics of distribution far enough to get partway down the potential Tail. The rest is left to the even more efficient economics of pure digital distribution. Both are Long Tails, but one is potentially longer than the other. (2006, p. 91)

In the absence of “pure digital distribution” one is left with the “tyranny of the shelf” (2004, p. 94). Anderson acknowledges that a distributed system of supply of physical goods can service long tail demand, and he uses Amazon as an example of a successful long tail business that relies on aggregated digital demand, but on distributed physical supply. He also points out, however, that Amazon has identified that their future lies in digitising their supply chain which has resulted in them investing heavily in print on demand technology.

These economics are potentially so efficient that they may someday make it possible to offer any book ever made. If you are a bookseller that means you won't have to be discriminating about what you do and do not carry in a print-on-demand edition, because the costs of making a mistake are also essentially zero. (2004, p.95)

It is apparent that library “goods”—at least in so far as they are stored by libraries and will potentially form repository collections—are persistently physical. To the extent that the library long tail has been digitised for immediate, low-cost transfer, it is in the form of e-periodicals, a market sector for which control resides largely with publishers and aggregators rather than with libraries.

This is also a reminder that libraries are engaged with the storage and distribution of two quite different log tail components, periodicals and books. Periodicals fit quite reasonably with Anderson's tests for long tail commercial success. Not only is supply aggregated in the form of large-scale databases of full text e-periodicals and commercial document delivery services, but there is also aggregation of demand provided by interdisciplinary indexing and abstracting services.

The situation with books, however, is very different. The overwhelming majority are not available for digital storage and distribution, and notwithstanding the ambitious plans of Google this situation is unlikely to change rapidly. Books also suffer, despite the development of union catalogues, from inadequate aggregation of demand. Most library catalogues provide very scant access to the intellectual content of books, and in so far as aggregated and filtered content information is available, it is through the websites of commercial booksellers such as Amazon rather than library catalogues.

In the case of both periodicals and books it is therefore commercial services rather than libraries that are at the forefront of creating the digitised discovery and supply

services that can maximise the value in the long tail. Which raises the question of whether libraries need to remain in the long tail “business” at all?

Tyranny of the shelf

And if, for the above reasons, Anderson’s version of the Long tail is not completely analogous with the research library long tail, what might his model nonetheless tell us about the nature of legacy print collections and how we should be planning for their future management? One aspect of Anderson’s analysis of retailing method is quite revealing of the circumstances facing libraries—his description of the “tyranny of the shelf”. Although Anderson acknowledges the way in which the supermarket shelf has been cleverly managed and exploited (“It is the very embodiment of capitalism evolved” (2006, p.152)), he also describes in detail its manifest limitations as a medium for storage, location and retrieval. In particular, he argues it is the very limited expandability of the shelf, coupled with its various ontological failings as a storage system, which have for so long denied consumers access to the benefits of long tail consumption. At the same time, however, libraries—free of the need to justify shelf space in economic terms—have been able to offer an illusion of an “infinite shelf” which in turn allowed them the luxury of storing long tail content.

This situation, however, is rapidly being reversed. The ability of publishers and other information retailers to use digital storage has for the first time allowed *them* a version of the infinite shelf. For libraries, however, the very finite capacity of their physical shelves has become all too apparent—and indeed it is a limitation that has been exposed by the advances in digital storage being made by retail enterprises. We are therefore witnessing the ways in which the infinite digital shelf is allowing publishers and other providers to adopt long tail retail strategies. It is now commercially viable for them to store, aggregate and market back issues of periodical literature in ventures such as J-Stor and Science Direct. Whereas previously libraries alone carried the responsibility for the research long tail, it is now being shared by information retailers in a manner which offers convenience and cost-effectiveness to libraries and their users. By utilising this newly developed retailer capacity for storage, libraries can continue to offer access to the long tail—for those elements of their collections which are available in digital form—without the need to suffer the tyranny of the shelf required by local storage.

Conclusion: the short tail

The inescapable conclusion is that at a time when retailers are exploiting the long tail, libraries should be thinking of their own future storage in terms of the head and a short tail. It is the commercialisation and approaching ubiquity of the digitised research long tail that will allow libraries to diminish their level of engagement with an aspect of their collections that has been becoming increasingly expensive and burdensome. It is difficult (impossible?) to find any compelling reasons why individual libraries should retain collections which are typically underutilised by the standards set for other long tail contexts; are often stored in the most cripplingly inefficient circumstances; and which suffer high levels of duplication with other collections which are equally inefficiently stored.

This brings me back to the question of repositories. While individual libraries might increasingly find it attractive to focus on the development and grooming of the head, this may not be the case for a system as a whole. Not only do some libraries have legislated obligations for the indefinite storage of certain materials, but there are also convincingly good reasons related to the nature of research for original items to be efficiently retrievable within a network of libraries. It is also unlikely that as a profession librarianship is ready to cede its curatorial function to commercial interests.

The repository therefore appeals as the one viable form of storage of the print research long tail. Although it does not totally eradicate the problems with shelf storage, it does allow forms of high-density storage which can correct the worst of the inefficiencies—single depth storage; limited heights to shelving; and each shelf at less than 100% capacity—associated with normal library storage. At the same time the repository also permits the development of the efficiencies with regard to supply that are so championed by Anderson and Dempsey. The implementation of shared repository storage will allow libraries to focus local print collections and services on the head, while retaining and indeed enhancing the benefits of long tail access.

Exactly how repositories should be scaled (for example; regional or national; research libraries only or all library types etc) and funded, and the exact nature of the policy environment in which they are to function, remain open questions. Indeed it is seemingly inevitable that the responses to these issues will be context dependent. But notwithstanding the differences discussed above between Anderson's long tail and its particular manifestation in research library collections, it would seem that the logic of the long tail remains highly relevant—even if the outcome is that libraries respond by implementing a short tail policy for local collections, supported by long tail access via commercial databases of digital collections and shared repositories for print materials. For most libraries, the days of the local long tail may be coming to an end.

Anderson, C. (2004). The Long tail. *Wired Magazine*, 12 (10). Retrieved March 12, 2007, from <http://www.wired.com/wired/archive/12.10/tail.html>

Anderson, C. (2006). *The Long tail: How endless choice is creating unlimited demand*. London: Random House.

Chems Consulting. (2005). *Optimising Storage and Access in UK Research Libraries: A Study for CURL and the British Library*. Retrieved March 12, 2007, from http://www.curl.ac.uk/about/documents/CURL_BLStorageReportFinal-endSept2005.pdf

Dempsey, L. (2006). Libraries and the Long tail: Some thoughts about libraries in a network age. *D-Lib Magazine*, 12 (4). Retrieved March 12, 2007, from <http://www.dlib.org/dlib/april06/dempsey/04dempsey.html>

- Evolution or Revolution: Debating the future of the UK's higher education information infrastructure: A conference. (2005). *New Review of Academic Librarianship*, 11 (2), 153-176.
- Henden, J. (2005). The Norwegian repository library. *Library Management*, 26 (1/2), 73-78. *Library Management*, 26 (1/2), 73-78.
- Kempf, K. (2005). Storage solutions in a co-operative library system: the case of Germany/Bavaria. *Library Management*, 26 (1/2), 79-88.
- Nicholson, C. M. (2005). CASS: A collaborative academic store for Scotland. *Library Management*, 26 (1/2), 32-41.
- O'Connor, S. (2004). Collaborative strategies for low-use research materials. *Library Collections, Acquisitions & Technical Services*, 28, 51-57.
- Payne, L. (2005). Depositories and repositories: changing models of library storage in the USA. *Library Management*, 26 (1/2), 10-17.
- Reilly, B. F. (2003). *Developing print repositories: Models for shared preservation and access*. Washington: Council on Library and Information Resources. Retrieved March 19, 2007, from <http://www.clir.org/pubs/reports/pub117/pub117.pdf>
- Saarti, J. (2005). From printed world to a digital environment: the role of repository libraries in a changing environment. *Library Management*, 26 (1/2), 26-31.
- Storey, T. (2005). The Long tail and libraries. *OCLC Newsletter*, 268. Retrieved March 15, 2007, from <http://www.oclc.org/news/publications/newsletters/oclc/2005/268/thelongtail.htm>
- Task Force on the Artifact in Library Collections. (2001). *The evidence in hand: Report of the task force on the artifact in library collections*. Washington: Council on Library and Information Resources Retrieved March 12, 2007, from <http://www.clir.org/pubs/reports/pub103/contents.html>
- Vattulainen, P. (2004). National repository initiatives in Europe. *Library Collections, Acquisitions, & Technical Services*, 28, 39-50.