Long Tail: An Emerging Technology for LIS in the Network Environment

Sanjay Kumar Singh   Prasanta Kumar Deka

Abstract

The enormous growth of the web, starting from zero to everything in the last decade, has forced us to believe in the impossible again. It might be the long-heralded great library of all knowledge really being within our grip? Today we need a building about the size of a small-town library to house 50 petabytes. With the advancement of tomorrow's technology, it will fit into our iPod or may be some new device, which can store large amount of data. When that happens, the library of all libraries will ride in our purse or every individual's pocket. If it doesn't plug directly into our brain with thin white cords… Converting inked letters into electronic dots that can be read on a screen is simply the first essential step in creating this new library. The real magic will come in the second act, as each word in each book is cross-linked, clustered, cited, extracted, indexed, analyzed, annotated, remixed, reassembled and woven deeper into the culture than ever before. In the new world of books, every bit informs another; every page reads all the other pages… This paper mainly discusses about the long tail as an emerging technology for libraries in the network environment.

Keywords: Long tail, Emerging Technology, Internet.

1.   Introduction

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 (i.e. inventories and catalogues) and the traded items themselves, have become digital rather than physical. 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 the items which will not find a commercially justifiable place in the physical stock (the head) of a high-street retailer, but which is nevertheless 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.

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 nevertheless 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 un- commercial, the little known and the out-of-print; and by meeting demand via inter-library loan, 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. 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).

2.   The Long Tail

The Internet has changed markets. In the physical world, the costs of distribution, retail and consumption mean that an item has to generate enough sales to justify its use of scarce shelf, theatre or spectrum space. This leads to a limit on what is available through physical outlets and a corresponding limit on the selection potential of users. At the same time, the demand for a particular product or service is limited by the size of the population to which the physical location is accessible.

Anderson provided some interesting numbers to show the impact of this phenomenon in his original article, and these have been updated on his website <http://www.thelongtail.com/>. He notes that the aggregation of the long tail is a major part of the business model of the leading Internet services (Amazon, eBay, Google, etc.). Google, for example, services the long tail of advertising – those for whom the bar was too high in earlier times of scarce column inches or broadcast minutes. And by aggregating demand, delivering a large volume of users, they increase the chances of the advertisement being seen by somebody to whom it is relevant. Of course, merely being on the web is only a part of the issue. What the web allows is consolidation. Anderson's examples are massive, consolidated web presences. This consolidation has two aspects: aggregation of supply and aggregation of demand. All are important.

Among five things about the aggregation of supply and demand, the first is transaction costs, the costs incurred – whether in attention, money, expertise or some other resource – in achieving one's goal. High transaction costs inhibit use: they increase the friction in the system; low transaction costs encourage use: they increase flow through the system. iTunes for example, has low transaction costs. The burden of discovering tracks of interest, transacting for their use and downloading them is low. The tracks are immediately available. Netflix has higher transaction costs given the delays caused in the mail system, but still works to provide as frictionless a workflow as possible for the user. The second thing is the availability of consolidated data about choices and behaviors. Netflix, Amazon, Rhapsody and others refine their service based on what they know of their users' choices, mined directly from the aggregated click stream. This allows them to develop services that can further develop reflexively based on usage, and that can be tailored around particular behaviors and preferences. The third thing to consider is inventory. These large web presences consolidate inventory are not encumbered by the costs of massively redundant, just-in-case inventory, scattered through multiple physical delivery points. This consolidation may happen in virtue of the digital nature of the collections, as with iTunes. Or, where physical inventory is involved, as with Amazon, they can consolidate in strategic locations, or with particular suppliers, as inventory need not be tied to physical storefronts. And, of course, consolidation of inventory may reduce transaction costs by streamlining fulfillment. The fourth thing is about navigating the consolidated resource. Google introduced a major innovation with its ranking approach, by aggregating and mining the linking choices made by web page authors. Amazon is interested in rich interconnection through reviews, wish lists, reader selected lists, the various 'phrases' (capitalized and statistically improbable), and so on. Amazon provides a rich texture of suggestion. In each case, simple aggregation is not good enough: also needed are effective ranking, recommending, and relating. And finally, large web presences help aggregate demand. The level of use of a resource partly depends on the size of the population to which it is accessible. One aspect of the long tail argument is that the aggregation of demand – extending the population to which a resource is accessible – means that resources have a better chance of finding interested users. In other words, use will extend down the long tail.

3. Libraries and the Long Tail

For libraries focus may be given on the issues like the aggregation of supply, and the aggregation of demand from reader point of view. For convenience primarily focus has been given on books, drawing in other resources occasionally.

3.1. Supply and Demand in Libraries

Libraries have been subject to the same physical constraints as bookstores, even though within a different service context. The library collection is not limited to the current or the popular, the library has some responsibility to the historical record, to the full range of what has been made available as

well as to what is now available. This type of responsibility varies from library to library. The library has met that responsibility in two ways by assembling a local collection, and by participating in systems of extra- and inter-library provision. These latter systems may be organized in different ways; the resource-sharing consortium is a common pattern, and a library may belong to many. The library collection is driven by local perception of need and available resources, collection development activities exist to balance resource and need. A large research library and a busy public library system will have different profiles, but both are influenced by physical constraint. In this materialistic world, the transaction costs of access to a distributed library collection are high, so those libraries that could afford it sought to accumulate large local collections in order to aggregate supply locally. Think of the large library collections. And, indeed, we are still measuring research library strengths by number of volumes. A busy public library may move towards the bookstore form.

3.2.   Transaction Costs

A library user has a range of discovery tools and services that provide access to a fuller range of scholarly and learning materials. A well-developed apparatus of deposit libraries, resource sharing systems, union catalogs, cooperative collection development, document supply, and other collaborative and commercial services, in turn supports this. This apparatus may be imperfectly and intermittently articulated, but it is a significant achievement on the other hand. What an individual library may not be able to supply should be available within the overall system in which libraries participate. However, this availability is bought at the expense of some complexity, which in turn means that the transaction costs of using the system are high enough that some needs go unrecognized. A library user may not be familiar with available tools or may not be aware that other materials are available. Local policies may restrict some types of access. Thus, historically, one can say that while library services explicitly aim to aggregate supply and demand both to meet user needs and to maximize use of resources within an overall apparatus of provision imperfect articulation of that apparatus means that users are variably served. To make this more concrete think about the D2D chain, discover, locate, request, and deliver. Here lack of integration increases transaction costs. By integration, there are many discovery options and between processes, the processes are not always connected in well-seamed ways.

◆ **Discover:** The discovery experience is a fragmented one. A user has a range of discovery tools available and may not always know which is the most suitable one. This is especially the case with the periodical literature, in which case the deployment of meta search approaches is a partial response. Even for books, users may have to navigate a patchwork of catalogs to find what they are looking for; search costs are high. What might one do? One approach is consolidation: fewer but larger pools of metadata to support discovery would help. Another is syndication, moving the metadata to where it might more readily rendezvous with the reader. Syndication is used as a general term to include such ideas as letting metadata flow into citation managers, search engines and other resources, and to expose it in services upon

which other applications may build. The latter is familiar from Amazon, which can make its data and services available in other interfaces through its APIs.

◆ **Locate:** Having identified an item of interest, a user needs to find a service that will supply it. This may be as simple as noting a call number and walking to a shelf. Or it may involve a resolution service that actually provides several service options. Or it may involve a further discovery experience in a library resource if the item was originally found outside the library. This latter case is especially interesting, as library users have many more discovery options outside the library than within it. What is needed is a way of connecting the discovery experience to a library service.

◆ **Request:** This is another transaction, which may involve one or more steps. It can be simple, as in placing a hold, or more complex if a form has to be filled out, and so on. Increasingly, libraries may want to route requests in several directions, allowing a user to buy from Amazon, initiate an ILL request, initiate a document supply request, or place a hold on the requested material.

◆ **Deliver:** Again, several potential options exist for resource delivery, which can involve more or less difficulty depending on how the delivery options are presented and on the disposition of supplier and user. This ties interestingly to the inventory question. At every stage, there are potentially many processes that need to be connected, and they potentially need to be connected to each other in different combinations. The better connected, the lower the transaction costs. Indeed, it is interesting to wonder if resolution services will move more to the center of library operations, as they are effectively service routers connecting multiple discovery experiences to multiple fulfillment services.

4.   Choice and Behaviors

Transactional and behavioral data is used to adapt and improve systems. In the library community these opportunities are not yet fully exploited. Examples of such data are holdings data (choices made by libraries), circulation and ILL data (choices made by users), and database usage data (choices made by users). Libraries are increasingly interested in using this data to refine services and build new services. Think of recommender services based on circulation data, for example. As new services and user behaviors co-evolve in changing digital spaces, it is likely to capture new forms of data.

4.1   Inventory

The historic library model has been physical distribution of materials to multiple locations so that the materials can be close to the point of need. And again, in the network environment, of course, this

model changes. Resources do not need to be distributed in advance of need; they can be held in consolidated stores, which, even with replication, do not require the physical buildings. As we move forward, and as more materials are available electronically, we will see more interest in managing the print collection in a less costly way. There are two medium-term questions that are of great interest here. First, what future patterns of storage and delivery are optimal within a system (again, where a system may be a large library system, a state, a consortium, a country)? Think of arranging a system of repositories so that they are adjacent to good transport links for example, collectively contracting with a delivery provider, and having better data intelligence for populating the repositories, based on patterns of use and demand. Second, think of preservation. What about the long-term costs of print preservation? Long-term may see a shift of cost from print to digital, but this can only be done if the costs of managing print can be reduced, which in turn means some consolidation of print collections.

4.2 Navigation

Library aggregations have not exploited the structure of the data very effectively to support navigation. The interest in faceted browse, FRBR (Functional Requirements for Bibliographic Records), recommendation, ranking by holdings or other data, and so on is testament to a realization that better ways to exploit the large bibliographic resource are needed. Ranking, recommending, and relating help connect readers to relevant materials and also help connect the more heavily used materials to potentially useful, but less used, materials.

4.3 Aggregation of Demand

The library resource is fragmented. It is fragmented within the library (there are many databases to choose from; they may be organized in a different ways in different libraries). In the new network environment, this fragmentation reduces gravitational pull. It means that others to whom, nevertheless, the resources are potentially useful may not reach the persistent or knowledgeable user prospects resources, but them. Additionally, the library resource cannot be very well assimilated into user workflows. The availability of RSS feeds, APIs, and other approaches are making it possible to insert the library into the user environment (rather than always expecting the user to come to the library environment). There are two issues here. The first is that libraries may need to do more work to aggregate demand within their own institutions. And one approach to this is to consolidate the library web presence (eg. meta search) and to project library services into user workspaces (eg. embedding database searches in course pages). The second issue is that it may be difficult for individual libraries to aggregate demand above the individual library level. Union catalogs and resource sharing systems have historically operated above an individual library level, and it is visible that organizations supply those services thinking about how to re-develop as major web presences that help aggregate demand (backed up by aggregated supply). Library organizations are also very keen

to be visible within the major web-based search engines and book selling sites. Of course, one way for a library to try to reach its local audience is to make its resources visible in these major web presences, which is where its users spend much of their time and attention.

This provides an interesting perspective from which to view Google Scholar and Google Book Search, in particular their interaction with libraries. Take Google Book Search, what Google is doing here is potentially aggregating demand for books and it will be interesting to see what influence this has on their use. Presumably a case has been made that there is potential interest in the full scope of those collections, or, in other words, in moving down the long bibliographic tail. They are also aggregating demand for books and journals through Google Scholar. And, to avoid frustrating users, they are aggregating supply behind the discovery experience. In addition, they are working with OCLC to connect the Google Scholar discovery experience to the Find in a Library option for fulfillment. What OCLC is doing is making metadata about those books available to the major search engines and routing users back to library services, to complete the D2D chain for books. To the extent that a large amount of materials are made available through these services, Google is aggregating demand, aggregating supply, and reducing transaction costs.

5. Logistics and Libraries

Libraries have rich deep collections, and the aggregate library system is a major achievement. However, in current network environment, libraries compete for scarce attention. This suggests that if the library long tail is to be effectively prospected then the cost of discovering and using library collections and services needs to be as low as possible. This is a logistics issue. Logistics is about matching supply and demand in a timely fashion across a network of potentially many parties. Within a particular domain, this is what libraries have always done, and some of the recent innovation in libraries has been precisely to automate supply chains. A few ways of improving aggregation of supply and demand are given below:

- ◆ Unify discovery experiences: Fragmentation is costly, and fewer but larger resources might help.

- ◆ Project library discovery experience into other environments: search engines, browser tools, RSS aggregators, etc.

- ◆ Better integrate D2D, both within operation and between operations: The aim should be to be able to place a get it button anywhere and guide the user through simple choices.

- ◆ In the medium term, explore how inventory and distribution are managed across a system: This should be done whether a system is a library, a consortium, a state, or a country.

- ◆ Utilize better intelligence within the network: This involves better representing the entities within the network.

◆ Provide transaction support: In an environment of multiple transactions between libraries it is useful to have a way of tracking and reconciling between libraries.

◆ Aggregate demand through significant web presences: If more users are exposed to library collections, the collections will be used more.

6. Conclusion

A fast transformation mainly in the hi-tech area of the unrestricted globe, are more and more philosophical and more frequent than at any other time in this ever-changing history. Whatever profession we hold, the day-to-day reality of our workplaces is change. A library is in the midst of change on a groundbreaking and transformational level. The biggest challenges in library profession today are changes like continuous change, technological change, copyright and ownership change. Even the tides of social and political changes in India affect libraries very seriously. To prosper as library personnel involves constant learning and relearning of theories, technologies, trends and patterns in the library profession. It is never wise to be self-satisfied and accept that what we are doing and thinking today will really apply tomorrow. As library personnel, we could all do the profession a favour by reminding everyone more often about the good that libraries do. As library professional, we should always make an effort to become very well versed in new emerging technology as it relates to our area and feel confident to enlighten our users with our knowledge.

Science and technology multiply around us in the same speed as the population and information are increasing like a jet plane. To an increasing extent, they dictate the languages in which we speak and think. Either we use these languages, or we remain mute. It is certainly important to be able to define and understand all the emerging technological and library lingo, it is even more important to be able to determine what is relevant to our own learning process, and relevant to getting the job done the best way. It is easy to become inundated and overwhelmed with data and forgets to see the library for the books. The libraries have many books that there is no longer enough room for reader. Libraries collectively manage a long tail of research, learning and cultural materials. However, we need to do more work to make sure that this long tail is directly available to improve the work and lives of our users.

References

1. Anderson, Chris. The Long tail: How endless choice is creating unlimited demand. London: Random House, 2006.
2. Anderson, Chris. The long tail. Wired Magazine, 2004, 12(10).
3. Dempsey, L. Libraries and the Long tail: Some thoughts about libraries in a network age. D-Lib Magazine. 2006, 12 (4).
4. Dempsey, L. Weblog. Libraries, Logistics and the Long Tail. February 15, 2006.

5. Storey, T. The Long tail and libraries. OCLC Newsletter. 2005, 268.

Webliography

1. http://libres.curtin.edu.au/
2. http://orweblog.oclc.org/archives/000949.html
3. http://orweblog.oclc.org/archives/000955.html
4. http://www.dlib.org/dlib/april06/dempsey/04dempsey.html
5. http://www.oclc.org/news/publications/newsletters/oclc/2005/268/thelongtail.htm
6. http://www.oclc.org/resourcesharing/features/feemanagement/default.htm
7. http://www.thelongtail.com/>.
8. http://www.wired.com/wired/archive/12.10/tail.html

About Authors

Dr. Sanjay Kumar Singh, Lecturer, Department of Library and Information Science, Gauhati University, Guwahati-781014, Assam.
E-mail: sksgu1@gmail.com
Mr. Prasanta Kumar Deka, Librarian, K. C. Das Commerce College at Guwahati. Assam.
E-mail: dekapk@sify.com