Introduction


> Hitherto library services have been dominated by local catalogs, local collections, and great inequalities in the geographical distribution of services. The constraints on library service are changing right now… All of this requires us to think again about the mission of the library, the role of the library, and the means of providing service. For the first time in one hundred year we face the grand and difficult challenge of redesigning library service.¹

Today academic Libraries are in the midst of a fundamental transformation of the basic strategies they use to provide documents and the information they contain. In doing so they are altering centuries old practice. We are finally getting around to the work Buckland charged us to begin twenty years ago.

These new strategies are made possible by transformations in scholarly communication that in turn has been made possible by the advent of the Internet and related technologies. As Clay Shirky has put it, “The moment we are living through, the moment our historical generation is living through, is the largest increase in expressive capacity in human history.”² Ray Kurzweil has clearly documented that when something becomes digital the pace of change becomes exponential.³ Scholarly communication has become digital and the pace of change will not stop or slow anytime soon.
I have laid out my views on how libraries will need to adapt in my 2007 article “A Strategy for Academic Libraries in the First Quarter of the 21st Century.” My more recent 2013 article “From Stacks to the Web: the Transformation of Academic Library Collecting,” which focuses on the changes in collecting. It is my belief that academic libraries will of necessity change their collecting practice even though such changes will be anathema to some and disruptive to many. They will do so because in the end the result will be a product that is cheaper, faster, and easier. In the end students, scholars, and the reading public will be better served.

Scholarly communication has been in a period of transformation for at least the past 40 years. Initially the change was driven by economics not technology. Academic libraries have largely been in a reactive mode responding to the changes in scholarly publishing. In particularly they have been struggling to respond to the takeover of scientific, medical, and technical journal publishing by a small number of for-profit publishers who realized that scientific journals were monopoly goods, acquired the most highly ranked titles, and then exploited their monopolies at the expense of libraries. This process began while journals were still distributed in print but has carried over to and been pressed even harder as the distribution became digital. Jean-Claude Guédon provides the best telling of this story in In Oldenburg’s Long Shadow: Librarians, Research Scientists, Publishers, and the Control of Scientific Publishing. The academic library response to being exploited was reactive; it was not forward looking or strategic. And the truth was there was little libraries could do, aside from complaining, because they had little or no bargaining power. The big for-profit publishers held all of the cards and they played their hand well.

But as I outline in my articles, I there has been a shift. Academic libraries can now alter what they do in a strategic way. They can look forward and chart the course they wish without the constraints that previously held them back. The new digital technologies allow for new business models and radically lower the costs of entry into publishing. As Clay Shirky frankly puts it, “With the old economics destroyed, organizational forms perfected for industrial production have to be replaced with structures optimized for digital data. It makes increasingly less sense even to talk about a publishing industry, because the core problem publishing solves — the incredible difficulty, complexity, and expense of making something available to the public — has stopped being a problem.” The for-profit publisher’s monopolies still exist, at least for now, but they are less compelling and are quickly being eroded around the edges. There is good reason for hope.

I believe strongly that the libraries that move assertively and change their practice, particularly their collecting practice, will create a competitive advantage for their institutions. Those that choose to move slowly will be a drag on their campuses.
This paper will cover much of the same ground as my two *College & Research Libraries* articles, but will do so with the help of a series of eleven pictures. My hope is that the pictures will be worth some number of words and that they make the path forward clearer. I tend to be visual in my thinking and for me at least the seeing as well as reading makes understanding easier.

**In the Beginning**

As I sit at my computer and watch the words I type appear on the screen, it is hard to comprehend the change that has occurred in the last fifty years. Documents, for well over a thousand years almost always print on paper documents, have been the mechanism that moved our words and thoughts through space and time. Now these tangible physical objects have given way to bits. Bits are not quite physical things in the way print on paper is, though of coarse they are physical things after their own fashion. They are much more flexible and fungible. They can be copied and can travel very far very quickly at nearly no cost. This is the world we now live in.

It was of coarse, not always so. Since the invention of writing, physical objects were required for text to be kept and read. For the past several hundred years the dominant form of object used for this purpose was the codex printed book and its cousin the journal. The printed book is a remarkable technology. It is durable and easy to use. And they are often special to us in strange and powerful ways.

Probably for as long as there were documents, there were people and organizations that collected them. But libraries as we know them can into existence in the latter half of the Nineteenth Century as industrial technologies were applied to publishing and the production of books and journals. Their numbers increased as universitites expanded and the pace of discovery increased. Melvil Dewey and his colleagues invented new forms of library practice to respond to the expanded production of books and these practices made it possible for libraries to grow so that by the middle of the Twentieth Century academic library collections of millions of volumes were not uncommon.

As we consider the changing world of scholarly communication and the way academic libraries will respond to it, we need to begin with an understanding of where we have come from. We come from the world of industrial book production and the library practices developed in response. We come from the world of large localized print collections.

So what did this world look like?
Painting the Picture

Let’s begin by imagining the universe of scholarship. Think of it as a series of concentric circles as shown in the figure below.

In the inner most circle are the items that are most used. As a general rule these items are also the items that are most widely held in libraries and personal collections. Also as a general rule these items are more general in their scope and content, although we would find “classic” works in this inner circle as well. As we move to the outer circles the works are less used and less widely held. They also tend to be more specialized. At the far outer edge the items are unique, found in special collections and be definition are held in only one place.
So the structure of the universe of scholarship can be defined as in the second figure.

This “top down” view of the universe of scholarship does not tell the whole story.

The distribution of items in the universe of scholarship is not even. Librarians often speak of the distribution as 80/20 with the inner 20% of the universe providing 80% of the use. This general understanding was grounded in sound research which dates back decades, Bradford’s law of scattering for journals, was established in the 1930s, and while modified by later studies was found to generally sound. Allen Kent’s Pittsburgh comprehensive study for books was published in the late 1970s found the 80/20 rule to generally apply. Interestingly, a recent study by OCLC and OhioLINK of the holdings and circulation records of the OhioLINK libraries suggests that the when we look beyond a single institution the core is much smaller than had previously been thought. To quote the study
report, “The most fascinating result of the study was a test of the ‘80/20’ rule. Librarians have long espoused the belief that 80% of a library’s circulation is driven by approximately 20% of the collection. The analysis of a year’s statewide circulation statistics would indicate that 80% of the circulation is driven by just 6% of the collection.” The important implication of this work is that large portions of the collection are likely to generate little if any use.

For our purposes the precise numbers are less important than the general idea that there is a small core of items that account for the majority of use and then a very long tail of less used items. Thus our universe of scholarship would look something like the figure below if view in cross section.

To fill out the picture imagine that the different quadrants of the circle represent the various areas of knowledge. This is of course a grossly simplifying assumption as different disciplines clearly have different types of items in the universe of scholarship — more journals in the sciences and more books in the humanities. And the distribution of use might be expected to be somewhat different as well. But for our purposes the differences are less important than the overall picture that holds more or less across the whole.

So the let’s consider our picture of the universe of scholarship to be shown in the figure below.
What We Know about the Patterns of Book Use

Librarians, or at least those who paid attention to the research of the field, have understood a number of things about the use of library collections for a long time.

First, the past use of a title, though not a prefect predictor, is by far the best predictor of future use. Herman Fussler and Julian Simon conclusively documented this finding in their study in the 1960s. As they state, “A not unexpected though crucial finding was that past use over a sufficiently long period is an excellent and by far the best predictor of future use... The variable of use is sufficiently powerful that for libraries with 20-year use records the objective characteristics make little objective contribution.”
Beyond the value of past use as a predictor of future use, Michael Buckland points out two other well understood drivers of book use in his *Book Availability and the Library User* published in 1975, “There is empirical evidence that the use of documents tends to follow certain patterns. The two patterns that are most relevant to the problem of managing library stock in order to maintain book availability in relation to the patterns of demand are:

(a) The distribution of demand over titles in a pattern usually known as Bradford’s Law of scattering;

(b) The distribution of demand over time in a well-established pattern of obsolescence.”

So we know that a small portion of collections generate most of the use and that new and previously used titles are likely to be contained in that core.

**Library Collecting in the Late Print Era**

So what did librarians do with this understanding of the nature of book use? Daniel Gore put forward a well-reasoned argument that this research argued for an actively managed library collection that would maximize the use of library space. In practice Gore’s proposal and the research on book use in general was completely ignored.

Librarians did not develop collections with the goal of maximizing use or the return on investment, rather they build collections to include as much of the long tail, or the outer circles in our model, as they could afford.

There were several justifications for this strategy. First it could be argued that the purchase of items, even those that were unlikely to ever be used, was an insurance policy. Books once out of print where difficult and expense to acquire and interlibrary loan was slow and unreliable before the development of the OCLC interlibrary loan system in the late 1970s. Thus the argument for a “just in case” collection development strategy made some sense. This strategy was also practical from a political perspective. Faculty, especially at research institutions, viewed the library collection as a marker of institutional prestige. Actual use of the collection was less important than its existence. A larger library collection was better because it was a demonstrated faculty and departmental importance. The use of the ARL Index as the means of establishing the pecking order of research libraries made this clear. Presidents and trustees paid attention to the ARL index and so it mattered.

If we overlay a library’s collection on our universe of scholarship it would look like the figure below. Institutions that could spend more on libraries had bigger blue circles and poorer institutions had smaller collections that would be represented as smaller blue circles. Tellingly many libraries had policies that discouraged duplication of even high use titles even when they would have generated more
use. Collections at various institutions might lean toward one or more quadrants depending on the importance of the various areas of knowledge at a particular institution.

In reality collecting at most institutions was more complex. Most had areas of special interest and they collected at a very extensive level in some cases attempting to collect everything from the most common and highly used to the unique and seldom if ever used. This is shown in our picture by the wedge of blue in the figure below. It was also often the case that many libraries had acquired and in some cases continued to build collections of special and unique materials. These were often the result of a gift or the interest of a single faculty
member. Depending on the circumstances of the institution these collections might receive a great deal of attention even if they were only tangentially related to the institution’s mission. Or, they might be neglected because the institution had limited resources even if the material was unique and valuable.

The resulting collections of a typical institution would then look like the figure below.
What has Changed?

In the past 25 years, digital technologies have been applied to scholarly communication, particularly to book and journal publication and distribution in a variety of ways that now provide libraries the opportunity to alter their core strategies.

The overall change is obvious: scholarship is now digital. Almost without exception all scholarly works begin life as bits, somewhere in someone’s word processor. The work may or may not be printed on paper at various stages of its life, though this is increasingly less often the case, and it is not a requirement for distribution, storage, preservation, or use. In fact, bits have many advantages over print on paper though we often seem unprepared to recognize these advantages and build them into practice.

Let’s look at some of these advantages and how they might be applied.

One: The print book market has become really efficient. Digital technology and worldwide communications make markets more efficient. This is especially true markets for items like books where a very large number of very specialized items are being pursued by a very small number of geographically dispersed parties with very specialized interests. In the world before markets powered by digital technologies you needed lots of bookstores, both for new and used books, and lots of middlemen. When you have a digitally powered marketplace all you need is eBay or Amazon. Chris Anderson’s example from his article “The Long Tail” clearly demonstrates how things have changed. As he puts it, “If the Amazon statistics are any guide, the market for books that are not even sold in the average bookstore is larger than the market for those that are.” For libraries this development is important because it is easy, quick, and generally inexpensive to acquire almost any print book regardless of whether it is in or out of print.

Two: The supply chain for print books is increasingly digital. It is now common, especially for books with limited sales potential, for books not to be printed and warehoused. Rather, they are only printed on demand. This is more efficient and it means that books no longer go out of print. This makes it possible for libraries to delay purchasing with a minimum of risk.

Three: E-books and e-book readers are becoming widely used. For many people e-books have become an acceptable and often a favored way of reading. For libraries the most important thing about e-books (and other e-content) is that, at least in theory, it can be purchased and delivered to the user in real time. This means that libraries do not need an inventory of pre-purchased items to deliver content. This makes possible user or patron driven purchase plans where by definition all the books that are acquired are used. Librarians have concerns about such plans, and they should as these plans are more effective than
librarians are at identifying what library users want, as Lynn Sutton puts it, “Professional collection development librarians have limited success in projecting the interests and demands of their users as measured by collection usage.” Selecting books has long been a core part of professional practice, so it will be hard to give up, even if it is shown to be inefficient and wasteful.

Four: The article is becoming disaggregated from the journal. As with music, digital delivery means you only need to buy the track or article you want, not the whole album or journal volume. Costs can be as much as $50 per article from commercial suppliers, but given the subscription cost of some journals, this can be a very good deal. Services like the Copyright Clearance Center’s “Get It Now” make the process easy across large bodies of content. It can probably be configured so that users would not be able to tell whether the article came from a library subscription or was purchased as a separate item. For libraries this means the question as to whether or not to subscribe to a journal is not, “Is this an important title?” Rather it is, “Which is cheaper, a subscription or article-by-article purchasing?”

Five: New publishing business models are now possible. The most important new model is open access that will be discussed at length below, but others may be possible. The key is as highlighted in the Clay Shirky quote in the introduction, the cost of publishing, at least for certain modes of publishing, has declined significantly. As is often the case with change of this magnitude, established organizations are having trouble adjusting their practice to take advantage of the available new economics, but they exist and the disruption that will result will create new products that over time.

What This Means

I believe there are two fundamental changes that we can anticipate that will allow libraries to alter their approaches to collection building.

One: Inventories are no longer required. While it is not yet quite true, we can begin to think of the universe of scholarship not as something that has to be collected and stored, at least not by individual libraries at the local level. Libraries do not, or soon will not, need inventories to provide documents to their users. There may be situations where inventories make sense, but this is likely the exception rather than the rule. The question then becomes what is the most cost effective way to provide the most content.

I believe that there will be circumstances where subscriptions will be the most cost effective solution and other circumstances were purchase-on-demand will be more cost effective. What will not be an efficient solution will be what has been established practice, item-by-item selection by librarians in anticipation of use.
Two: Open access is inevitable and will become the business model for much of scholarly publishing. I have made the case that open access publishing is a disruptive innovation as the business theorist Clayton Christensen would define it, and that as such it will inevitably become the dominant business model for scholarly publishing. As Christensen points out disruption innovations tend to impact not only a set of producers, rather disruptive innovations impact the full value chain of an industry. Academic libraries are part of the value chain of scholarly publishing and as such we are likely to see our business models at risk in the coming disruption to nearly the same extent that scholarly publishers will.

Let us look at how each of these fundamental changes will affect how academic libraries will begin to think about collections.

Inventories are No Longer Necessary

If we return to our picture of the universe of scholarship we can begin to imagine a world in which the library can purchase in the traditional manner a much smaller core of the universe. It is also likely to continue to acquire extensively in areas of special interest and to maintain special collections. But importantly, now it can view the whole of the universe as part of its collection. It may not yet actually be purchased, but it could be quickly and easily in response to a request from a user.

The result is a view of the library collection from the users perspective that includes everything in the scholarly universe. In fact, only the richest of libraries will be able to make this claim in this extreme a way, but many libraries will be able to define the piece of the university they will make available to users and this will be significantly larger than their current collections.

In some ways most academic libraries do this already with their interlibrary loan operations. One way to think about what will be happening is that interlibrary loan will get so good that an increasingly large portion of user needs will be met in this way. As this happens interlibrary loan will increasingly be a misnomer. What libraries will actually be doing is document delivery. They will be purchasing items from publishers or their agents rather than borrowing them from other libraries.

Interestingly Theodore H. Nelson, the inventor of hypertext, predicted this state of affairs 30 years ago. In a Publishers Weekly article in November 1984 he wrote, “There is no longer any question that publishing of the future — whether of words, pictures or musical notations — will be based on electronic information storage. Material will be shipped out an instant after a user’s request. The user will pay, and the publisher will be paid, at the moment of delivery. Publication will be a
system of storage and transmission on demand, and a newly published document becomes at once universally and instantaneously available to all those who have computers."\(^{18}\)

So from the user’s perspective the library’s collection looks like the figure below. The user’s library may have only purchased the dark blue portion of the scholarly universe, but what the user sees is that the entire universe is available. And it is.

From the library perspective the view is more like the one in the following figure. The dark blue is purchased in anticipation of use. This is a small portion of the scholarly universe. The dark dots are individual items that the library purchases in response to a particular user’s need.
Open access is often viewed as a cause or a political movement. It is seen as a way to strike back at the evil monopolists publishers who have captured scholarship and have exploited libraries for decades. This call to arms might provide a feeling of righteousness, but it confuses what is really going on. Open access is not a cause; rather, it is a new business model. It uses digital technology and the distribution channel of the Web, combined with a different economic structure, to publish scholarship in a different way.

As noted above, I have argued that this new business model is a disruptive innovation as defined by the business theorist Clayton Christensen, and as such.
will make scholarly publishing faster, cheaper, and easier for the consumers of scholarship. I have also argued that the whole supply chain for scholarly publishing will likely be disrupted. This will include publishers, bookstores and distributors, P&T committees, and libraries.

As Peter Suber defines it, “Open-access (OA) literature is digital, online, free of charge, and free of most copyright and licensing restrictions. What makes it possible is the internet and the consent of the author or copyright-holder.” Suber is mostly right, what makes open access possible is the technology of the Internet where the marginal cost of the distribution of copies of publications is for all practical purposes zero. It also requires the rights holder to be willing to be willing to give away at least some rights to their works, but importantly beyond this there is a need an publishing organization that can create a flow of funds that is sufficient to cover first copy costs and to maintain a distribution and, one hopes, a preservation, infrastructure.

Suber and other open access advocates generally think of open access as a mechanism to distribute scholarly journal articles. The funding model here is generally either article processing fees paid by authors, their institutions, or their grants, or institutional subsidies provided to journals by hosting institutions or others.

But if we take the new business model view of open access we see that other information products are open access. Wikipedia is an open access encyclopedia paid for by the donated time of contributors and philanthropic support. American Memory is an open access collection of digitized artifacts paid for by tax dollars channeled through the Library of Congress. The HathiTrust is a collection of digital books paid for by a collection of research universities. Google is a web index and search tool paid for by advertising.

We have begun to see that web scale allows for changes beyond the basic publishing model. PLoS One has created a mega journal where the article is the unit that matters, not the journal that contains it. In addition, PLoS One has cut the review cycle to a fraction of the time required by most traditional journals. Wikipedia has harnessed social production and coordinated volunteer labor to an unprecedented degree to create a huge repository of knowledge.

One of the key advantages all open access projects have is that they can dispense with the costs of customers. Or, as is the case when article processing fees are the funding mechanism, the customer becomes the author not the reader or the reader’s library. The down side of course is that you have to find another income stream. Accomplishing this is the riddle all successful open access projects have to solve, but there are many options and once one has been found there is a clear cost advantage over traditionally publishers.
The result to date is that some important parts of the scholarly universe are now open access. Interestingly, there are open access pockets all over the place. Some like Wikipedia are at the center — general in coverage and heavily used. Other are more specialized and less frequently used.

From the library perspective open access means two things. The first is easy to appreciate — if something is open access libraries don’t have to buy it. This is a relief on library budgets. The second is less appreciated, but maybe more important. Libraries are not required for readers to use the content.

Today the scholarly universe looks something like the figure below. Note that there are more dark blue circles on the outside of the universe. This represents new collections that a library will take on, usually as open access projects. Most often this will be content created at the institution or of strong interest to it. The library will take on the responsibility of curating this content for the world. If we think of this in the context of the diagram, the blue circles created by one library are new green circles for everyone else.
Academic Library Collecting in 2025

If my predictions for open access are correct, the scholarly universe will increasingly look like the figure below. Large portions, though not all of the scholarly universe will be open access. Some part of the core will still need to be purchased and there will still be some cases where purchase on demand will be required.

The net result should be lower overall costs to universities to provide access to the traditional scholarly universe. These savings will result first because purchase on demand will replace inventories as the means of providing readers
with documents. What this really means is that space formerly used for stacks
can be repurposed. And second because open access will make large parts of
the scholarly universe freely available. There will be some institutional costs for
open access and how these costs are distributed is still very much up in the air,
but for the system as a whole they should be cheaper than current practice.

That’s the good news.

The bad news is that this is not the end of the story.

**New Responsibilities**

If all we had to worry about going forward are things we have had to worry about
in the past life would be good. Unfortunately there is lots of new stuff that has or
will shortly become our problem. When information becomes digital it expands
because of several aspects of the nature of digital documents. First, digital
documents are easy to modify. Therefore versions are easy, as are mashups
and other forms of combination and reuse. Second, new digital forms are useful
in the scholarly conversation and as such they become part of the scholarly
record. Blogs, tweets, and the little are less formal harder to capture and
preserve, but many are valuable non-the-less. So the universe that we now
responsible for begins to look like the figure below, though the new pink circle
probably underestimates the size of the problem.

How this new content will be managed is unclear. Whether it will fall to local
libraries of web scale institutions is unclear. What is clear is that current library
practice is not up to the task.

It is likely that whatever savings accrue from begin able to more efficiently
managing our traditional responsibilities will need to be invested in this new task.
The Library as Information Subsidy

Finally, there is an interesting change of perspective that needs to take place with this change in the mechanism of providing documents to users. We are used to thinking of a library as a common good that an organization or a community provides to its members. Communities and organizations do this because it is understood, though not always at a conscious level, that individuals left to their own devices, will not seek or acquire all of the information they need to be
productive and successful. It is therefore in the interest of the community or organization to invest in the library as a public good to make acquiring information cheap and easy. In the print era the mechanism for providing the subsidy was a building full of books. In the digital world and using the strategy I have outlined what we have is more like an open checkbook that will purchase the books an articles any library user wants. It is the difference between a public park and giving citizens free tickets to Disney Land whenever they need recreation. Or, like the difference between a soup kitchen and food stamps.

With an old fashion library, a park, or a soup kitchen the system for rationing is the physical constraints of the space. In traditional libraries demand was controlled by the physical nature of our collections and spaces. It took time and energy to get to the library and if the book was checked out you waited for it to be returned. We now need to find a way to ration the scare resource.

We need to wrap our heads around this change. It feels different, because it is. There are economic efficiencies with this different approach, but it still seems odd.

We need to understand that libraries are at their core a mechanism for providing an information subsidy. It is the subsidy, making information cheap and easy, that matters, not the particular mechanism for doing so.

What is most important now is that we find ways to create as much good scholarship as possible and to make this scholarship freely and easily available to the largest possible audience.
Notes

5 Lewis, David W. “From Stacks to the Web: the Transformation of Academic Library Collecting.” *College & Research Libraries* 74(2):159-176 March 2013. Available at: http://crl.acrl.org/content/74/2/159.full.pdf+html and http://hdl.handle.net/1805/3252
I have suggested, only partly in jest, that our library should load records in the catalog for all of the books Amazon sells (I’m guessing that is 20 million give or take) with a location of the library’s remote storage facility with a delivery time of 48 hours. Our library, of course, has no remote storage facility and we would simply purchase the book when needed. I have outlined a similar proposal, again only partly in jest, in David W. Lewis, “The User-Driven Purchase Give Away Library,” *EDUCAUSE Review* 45(5):10-11 September/October 2010. Available at:

http://www.educause.edu/EDUCAUSE+Review/EDUCAUSEReviewMagazineVolume45/TheUserDrivenPurchaseGiveawayL/213955 and  
http://hdl.handle.net/1805/2304


http://www.ala.org/acrl/sites/ala.org.acrl/files/content/conferences/pdf/lsutton.PDF and  
http://hdl.handle.net/1805/2304


http://crl.acrl.org/content/73/5/493.full.pdf+html and  
http://hdl.handle.net/1805/2929
