Collection Directions:

Some Reflections on the Future of Library Collections and Collecting

By Lorcan Dempsey, Constance Malpas and Brian Lavoie

OCLC® Research

Collection Directions: Some Reflections on the Future of Library Collections and Collecting

Lorcan Dempsey
Vice President, Research, and Chief Strategist

Constance Malpas
Research Scientist

Brian Lavoie
Research Scientist

Contents

Figures

Abstract

This article takes a broad view of the evolution of collecting behaviors in a network environment and suggests some future directions based on various simple models. The authors look at the changing dynamics of print collections, at the greater engagement with research and learning behaviors, and at trends in scholarly communication. The goal is to provide some context within which libraries can discuss changing patterns of investment across collection categories.

Introduction

As the network continues to reconfigure personal, business and institutional relationships, it is natural that we will see changes in how library collections are managed in coming years: changes in focus, boundaries, and value.

This article looks at ways in which approaches to collections, broadly understood, are changing in a network environment. The treatment is discursive and not at all comprehensive; it focuses on several areas that we think are especially interesting. We present some models that may facilitate discussion about future directions, organized as follows:

- **The network context**: Libraries have been preoccupied with the shift from print to electronic materials, but changes in a digital network environment are further reaching than this, profoundly changing the nature of local library investments in collections and services, and reconfiguring how libraries coordinate with each other and with their providers.

- **The evolving scholarly record**: Libraries acquire, organize, and provide stewardship of the scholarly record. Ongoing redefinition of the scholarly record will drive changes in library and publishing practice.

- **The collections grid**: Libraries engage with different types of collections, which have different dynamics associated with them. Understanding the shift in the patterns of operational support for different types of resources is important to library planning and investment, both for individual libraries and for the networks of which they are a part.

- **The inside-out collection**: The dominant library model has been outside-in, where materials are purchased or licensed from external sources and made available to a local audience. The inside-out model, where institutional materials (digitized special collections, research and learning materials, researcher expertise profiles, etc.) are shared with an external audience requires new ways of thinking.

- **Managing shared print**: The print collection has been central to the identity of the library but is now on the threshold of major network reorganization. The emergence of

cooperative infrastructure, facilitated by the network, has enabled a transition from institutionally-organized stewardship toward group-scaled solutions.

- **Sourcing and scaling**: Collections will be managed at several levels, above the institution as well as within it. Choices about the optimum level (institutional, consortial/group, regional, global) for management are becoming more common, as are decisions about how to source activities (collaborative, buy from third party, etc.).

A Network Context

It is common to think about the changing position of the library in the broader network context in terms of the shift from print to digital, but this may be less important than three broad contexts we would like to introduce here.[1]

1. Unbundling and rebundling: transaction costs and system-wide reorganization.

2. A data driven environment: activities are becoming "informationalized," where more operations are automated and data drives decisions.

3. Research and learning behaviors are changing: libraries serve a constituency whose needs are also changing.

Transaction Costs: Bundles and Boundaries

Library services and organizations were formed in an era of physical distribution and interaction. The digital network reduces transaction costs, potentially changing the patterns of distribution and interaction. Transaction costs are those incurred in the interaction between organizations—the effort, time, or money expended in interaction with others. Although we do not usually think about it in this way, changing transaction costs are actually a major driver of library development. This may be a surprising claim, so we provide a brief review here.

Changing transaction costs are actually a major driver of library development.

The economist Ronald Coase famously argued that an organization's boundaries are determined by transaction costs—if it is more expensive to transact for a service in the market than to provide it internally, then it will be internalized. For example, at one time it was economical for an organization to manage its own payroll. However, now, many organizations have unbundled that functionality and contract for it in the marketplace. Lower transaction costs, driven by the network, have greatly enhanced the ability to

unbundle particular functions and source them externally. This dynamic has also facilitated the emergence of complementary, specialist providers who can achieve economies of scale by supplying multiple organizations with a particular service. In this way, whole industries have been reconfigured, as the physical distribution of functionality and expertise to multiple local sites is no longer always required. At the same time, consolidated platforms can concentrate functionality and data, and deliver the benefits widely. Think of the impact of Amazon on retail, or of Expedia on travel. Think of how UPS, ADP, Etsy or Square have allowed businesses to focus on what is distinctive to them, as they facilitate unbundling of local infrastructure to their shared platforms. Or think of how cloud providers (Amazon Web Services, Windows Azure, Rackspace, etc.) can accelerate organizational development by providing computing and applications capacity to startups and other organizations. As the need for physical distribution of expertise and materials diminishes, there is a trend to achieve economies of scale and greater impact by moving to network level hubs. The network favors scale in this way.

As the need for physical distribution of expertise and materials diminishes, there is a trend to achieve economies of scale and greater impact by moving to network-level hubs.

How does this relate to libraries? In a physical world, libraries assembled information materials close to their users. This gave rise to the model of the library that has dominated university perceptions until recently: that of a building which houses print collections and of an organization vertically integrated around the management of those collections. Each library deploys collections locally, as well as the systems and staff required to acquire, process, organize, and provide access to those collections. Preservation is a benign artifact of the print publishing model as materials are redundantly available across libraries.

In the print environment, it was convenient for each university to internalize a collection of locally assembled materials, to organize it, and to interpret it for its users. The alternative, where everybody was individually responsible for all of their information needs would be inefficient and expensive: the aggregate transaction costs across the university would be very high. These aggregate costs could be minimized by placing collections close to learners and researchers. This led to multiple local collections. It also meant that the bigger the local library was, the better it was seen to be, because it satisfied potentially more of local needs without having to go outside the institution.

As transaction costs go down in a network environment, it drives change across the system. Think of this from both infrastructure (supply side) and user (demand side) perspectives.

Infrastructure

There have been several waves of system-wide library reorganization, as activities previously a part of library infrastructure are now unbundled and sourced in consolidated platforms. Notably, these successively include the development of shared cataloging and resource sharing networks,[2] the move to a licensing model for the journal literature, and more recently for books, and the trend to cloud-sourced discovery and library management environments. Of course, the business arrangements in each of these cases is different, but they share the drive of reducing institutional system-wide transaction costs by unbundling institutional functions and consolidating them in shared network platforms. At the same time, negotiation and licensing moved partly into shared or consortial settings.

What this means is that libraries will increasingly collaborate around systems infrastructure (see for example, the growing interest in cloud-based shared management systems) and collections (see for example, the growing interest in shared print management arrangements), or unbundle these activities and externalize them to third parties where it makes sense (see for example, JSTOR, Portico, etc.). The reduced transaction costs of collaboration and externalization make this consolidation inevitable. Think of HathiTrust. A few years ago, it is likely that libraries would individually build infrastructure to manage digitized books and store them locally. Now a shared model is more compelling, as the network has reduced the transaction costs of creating and interacting with a consolidated resource. Of course, this makes the governance of the organizations to which these responsibilities are entrusted a critical community issue, another area requiring conscious coordination among stakeholders.

Library Users

On the user side, the change has been much more sudden and far-reaching. Whereas information creation and use may have been organized around the library, it is now coming to be organized around network level services which support individual workflows. For researchers and learners, the transaction costs of creating and using information resources have declined considerably. Access is no longer via a small number of physical gates, but has dispersed across many network resources. Think of this selection of very different services:

For researchers and learners, the transaction costs of creating and using information resources have declined considerably.

- arXiv, SSRN, RePEc, PubMed Central (disciplinary repositories that have become important discovery hubs);

- Google Scholar, Google Books, Amazon (ubiquitous discovery and fulfillment hubs);

- Mendeley, Citavi, ResearchGate (services for social discovery and scholarly reputation management);

- Goodreads, LibraryThing (social description/reading sites);

- Wikipedia, Yahoo Answers, Khan Academy (hubs for open research, reference, and teaching materials).

- GalaxyZoo, FigShare, OpenRefine (data storage and manipulation tools)

These network level services are important components of workflow and information use for researchers and learners. A large part of discovery activity has been unbundled to Google Scholar and Amazon.

Two related issues are worth noting here as we have seen a progressive decoupling of discovery and the collection.

The first is the relationship between what we might call the "available collection" (available to a library user because it is openly available on the web or because it is licensed by the library for the user) and the "global collection" (everything that is discoverable) from a discovery and fulfillment point of view. The available collection is a subset of the global collection. While the library historically provided access to the local or available collection, they now provide discovery access to more of the global collection (through the cloud-sourced discovery layers they are licensing). At the same time their users are using other resources to find materials (Google Scholar, Mendeley, etc.). The link between the available collection and the global collection may not always be apparent to a library user.

The second is that we have seen a progressive move away from purchasing and local storage at one end of a spectrum towards general facilitated access at the other. Examples of the latter are pointing users at Google Scholar, loading metadata for freely available e-books into the catalog, or creating resource guides which include freely available materials. In the middle is the licensing of electronic materials in packages the contents of which may change depending on the arrangement. This is a significant shift, as facilitated access becomes a service the library offers, which may or may not be attached to local materials.

As access and collections are decoupled in this way, it moves the library towards a set of services around creation, curation, and consumption of resources that are less anchored in a locally managed collection, and more driven by engagement with research and learning behaviors. At the same time it moves concern about preservation of resources away from the institution and towards a system-wide perspective, where incentives may be weak.

A Data-driven Environment

Manuel Castells uses the terms "informationalization" and "informational" on the models of "industrialization" and "industrial."[3] Informational activities are those where productivity is maximized through the use of knowledge, gathered and diffused through information technologies. "Informationalization" is visible at all levels. For example, supply chains, the disposition of goods around retail floors, or investment decisions are all increasingly influenced by behavioral data. Flows of people and materials follow the flows of data. An increasingly rich apparatus of instruments collects data about our activities, our environment and us.

> **An increasingly rich apparatus of instruments collects data about our activities, our environment and us.**

This will continue as larger parts of our domestic, business, and educational lives are being assisted by automated systems.

In the library context, we can see this in multiple ways. We are moving from a relatively static "document" based world to a more dynamic informational or data-driven one. Consider some examples:

- A **computational** approach is becoming more routine. Think of what is involved in managing repositories of digital materials, video recordings, and archives of web materials. For example, we will programmatically extract metadata from resources, as the volume of resources to be managed makes it difficult for manual processes alone to cope. We will mine text and data for patterns and relationships. In Franco Moretti's term, "distant reading" will complement close reading, as we look for patterns.[4]

- Resources are **social objects** that become nodes in a network environment. Think of "bibliographic" services such as Amazon, Goodreads, LibraryThing, WorldCat, and Mendeley. They each provide functional value: they get a job done; however they also provide network or social value as people make conversation and connections around resources of interest or importance to them. This in turn enhances the value of those services. Similarly, think of a reading list or a bibliography or a resource guide: they frame resources in the context of particular research or pedagogical interests. We need better ways of creating social value in library services.

- **Analytics** is now a major activity, as transaction or behavioral data is aggregated and mined for insight. We have become used to recommendations based on buying or navigation patterns. As more material is digital, as more business processes are automated, and as more activities shed usage data, organizations are manipulating

larger amounts of relatively unstructured data and extracting value from it. Within the library field, patterns of download, holdings, and query resolutions are being mined to improve services.

This trend has major implications for discovery, selection, acquisition, and management of collections. Consider the relative roles of DDA (demand driven acquisition) and library selected material, for example. Think of literature searching in an environment where researchers belong to several recommendations "networks" (e.g., Google Scholar, Mendeley, GoodReads, etc.).

Group or consortial environments are especially interesting in this regard, as the systems apparatus on which they run becomes more integrated and data-aware. Think of the data available to a group of libraries sharing interlibrary lending, acquisitions, discovery and DDA operations. We are looking towards an environment where this data will be used to trigger acquisitions, collection balancing between institutions, digitization, consolidation in shared print environments, disposal and so on. Analytics have become central, and the connections between usage, management, and purchasing/licensing decisions will become firmer as intelligent workflows are connected to networks of shared data about resources, usage and people.

Changing Patterns of Research and Learning

Libraries are not ends in themselves but serve the needs of the institutions of which they are a part. As those needs change, so do the requirements placed on the library.[5] Changes in the way research and learning are done are more important drivers of change than internal library developments.

Anticipating the discussion in the next section, we can see the scholarly record becoming the continuous digital recording of the research process, understood as a collective effort around the collection of data, documents and other resources, which will vary by discipline. The resources are collected, organized and kept in a repository that is shared with researchers active in that discipline. Research teams analyze the data via computational models, create visualizations to illustrate the modeled data, and enable other researchers to understand it. They adjust parameters, experiment with models, and annotate their observations of new results. The scientific knowledge is digitally recorded in, and dependent on, the complex infrastructures

> **The research process, its outcomes, and its aftermath will overlap, while traditional publishing of outcomes is simply one notification stream.**

where the research is done. Creating, fixing, collecting, and using the scholarly record happen continually, and the research process, its outcomes, and its aftermath will overlap, while traditional publishing of outcomes is simply one notification stream.

This is not the place to discuss such a big topic in detail. However, one important point connects to following sections. In a print world, the library's intersection with the scholarly life cycle was limited. The library collected outputs from scholarly activity and organized them as inputs to it. In a digital environment, the intersection points multiply, to include, potentially, support at all points in the lifecycle. This in turn drives a deeper engagement with the research and learning behaviors of the institution and individual researchers. Examples in a research context are the support for data curation, copyright, new forms of scholarly publishing/curation, bibliometrics and research profiling, data mining and visualization, and so on. In a learning context, support for research skills or curriculum development come to mind, as well as the types of support required for a range of new learning and teaching models. Consider the recent emphasis on the flipped classroom, online learning or MOOC developments, and the support requirements they raise.

A Network Context: Conscious Coordination of a New Relationship Architecture

The network is reconfiguring how libraries organize their systems and collections, and how faculty and students organize their research and learning workflows.

These trends demand service innovation but also institutional reinvention. Reconfiguring capacity across the network requires new "relationship architectures" that allow libraries to scale collaboration effectively.[6] Collaboration itself involves transaction costs, as time and effort are required to build trust and service networks across institutions. This means that libraries in well-established consortia (the Orbis Cascade Alliance, or OCUL in Ontario, for example) or in countries with public infrastructure (ABES in France, for example), may see those platforms grow to take on more responsibilities. At the same time, there will be a need for broader based shared capacity (to aggregate usage data for example, or to provide preservation infrastructure, or to provide shared data curation), again drawing on existing organizations (e.g., OCLC, Ithaka, HathiTrust) or developing new frameworks. This suggests that there will be a major need for much greater "conscious coordination" across the system to effectively leverage local investments into shared approaches which build capacity and impact.

The network is reconfiguring how libraries organize their systems and collections, and how faculty and students organize their research and learning workflows.

Lavoie et al for OCLC Research. 2014.
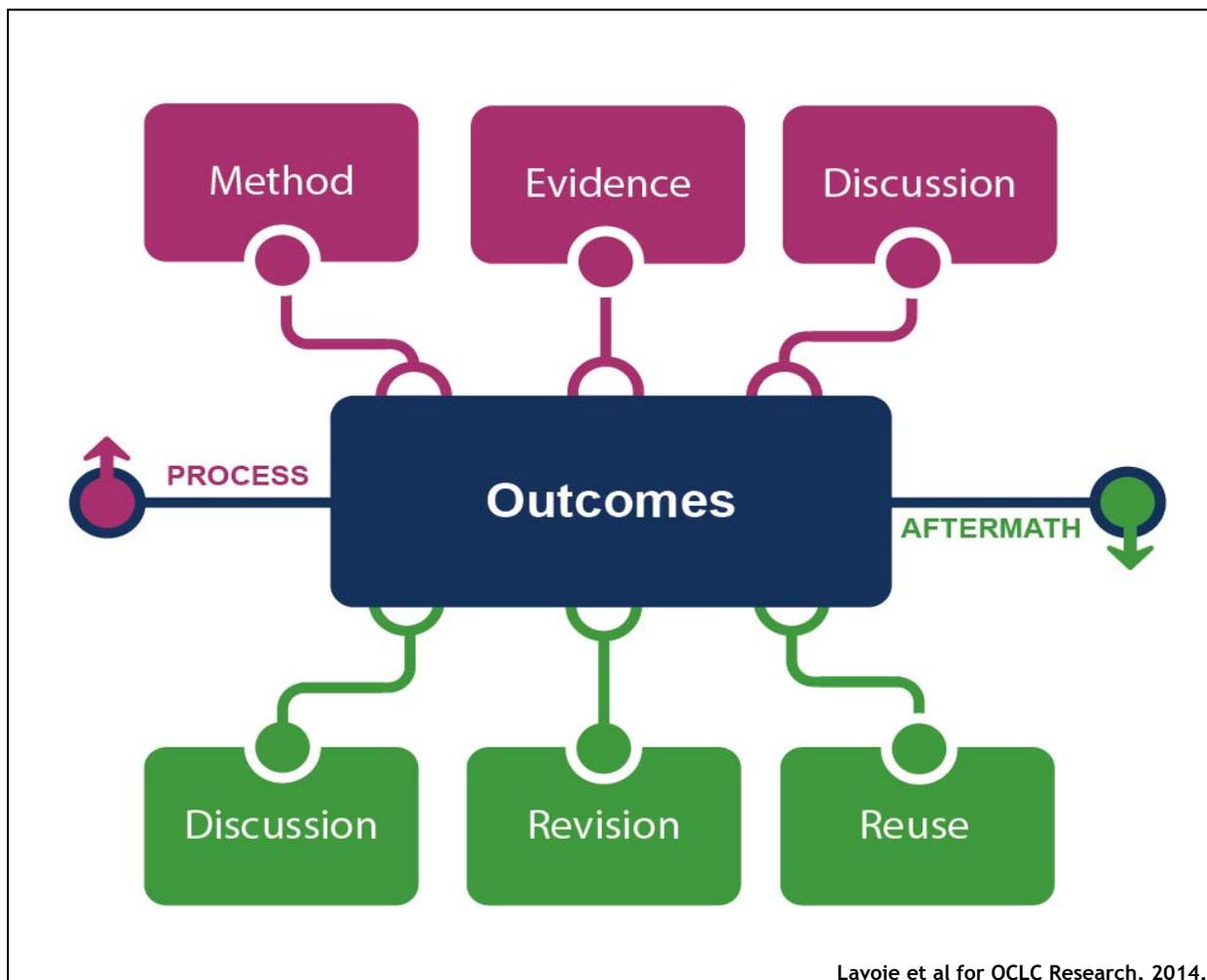
Figure 1. The evolving scholarly record

The Evolving Scholarly Record

The accompanying diagram (figure 1) conceptualizes future trends in the nature and scope of the evolving scholarly record.[7] The "final outcomes" of scholarly inquiry are shown at the center; while still consisting primarily of published text, these outcomes are increasingly supplemented by additional materials such as video, interactive programs, and complex visualizations. The rest of the scholarly record is divided into two broad areas: process and aftermath. The process phase refers to the process of scholarly inquiry, by which outcomes are produced. Within this phase, three categories of material are identified that could potentially migrate to the permanent scholarly record:

- **Method**: materials pertaining to methodological techniques and innovations (e.g., computer models, digital lab notebooks, sampling frames, experimental protocols)

- **Evidence**: the "raw materials" of, or inputs to, a research project (e.g., data sets, survey results, new or enhanced primary source documents)

- **Discussion**: formative discussions around scholarly ideas (e.g., preprints, listserv/blog discussions, conference presentations, grant proposals/reviews)

Once the outcomes from a research project have been formally published or otherwise made available, scholarly activities surrounding that piece of work may still continue in the aftermath phase, generating additional materials which potentially could be included as part of the permanent scholarly record:

- Discussion: discussions organized around published outcomes (e.g., through similar channels as those in the process phase, but also formal post-publication reviews and commentary)

- Revision: materials that build on, enhance, or improve published outcomes (e.g., the work may be enhanced with additional findings; errors may be corrected; clarifications made)

- Reuse: "repackaging" published outcomes for different venues or audiences (e.g., conference presentations, summaries, blog posts, versions for the "popular media")

Anchoring outcomes directly to the methods employed, evidence used, and formative discussions conducted during the process of scholarly inquiry helps contextualize and deepen our understanding of these outcomes, facilitate replicability, and leverage results into new research. Similarly, anchoring outcomes to the discussions, revisions, and reuse of the work that occurs after it is published or otherwise made available helps track the diffusion and evolution of ideas and findings as they circulate through the scholarly community and beyond.

Universities and libraries are interested in institutional and shared approaches to data curation, for example, as are some publishers and other service providers . . .

Systematically gathering and curating the range of materials produced during the process and aftermath phases of research, as well as published outcomes, results in a deeper and more complete record of scholarly inquiry. There are multiple stakeholders interested in constructing that more complete record. Universities and libraries are interested in institutional and shared approaches to data curation, for example, as are some publishers and other service providers, and interesting new services like FigShare have

emerged. Here again we see that "conscious coordination" will be needed as stewardship, discovery, and access of the scholarly record is increasingly distributed across multiple stakeholder communities: libraries, publishers, and other service providers.

Framing Collection Directions: The Collections Grid

The collections grid was developed a few years ago as a heuristic device to help think about the pattern of collection investments a library has. It simplifies and "flattens" some dimensions but has been shown to have utility. It organizes resources according to two values: "uniqueness" and "stewardship/scarcity." Resources that are unique, or rare, tend to be in one collection only. Resources that are not unique or rare tend to be in many collections. At the non-unique end of this spectrum are commodity materials, which are widely published or available through many channels.[8] Resources that are highly stewarded are things that attract library attention, have resources and time spent on them, have systems infrastructure devoted to them, and so on. Stewardship and scarcity tend to go together: we have developed stewardship models for materials that are relatively scarce.
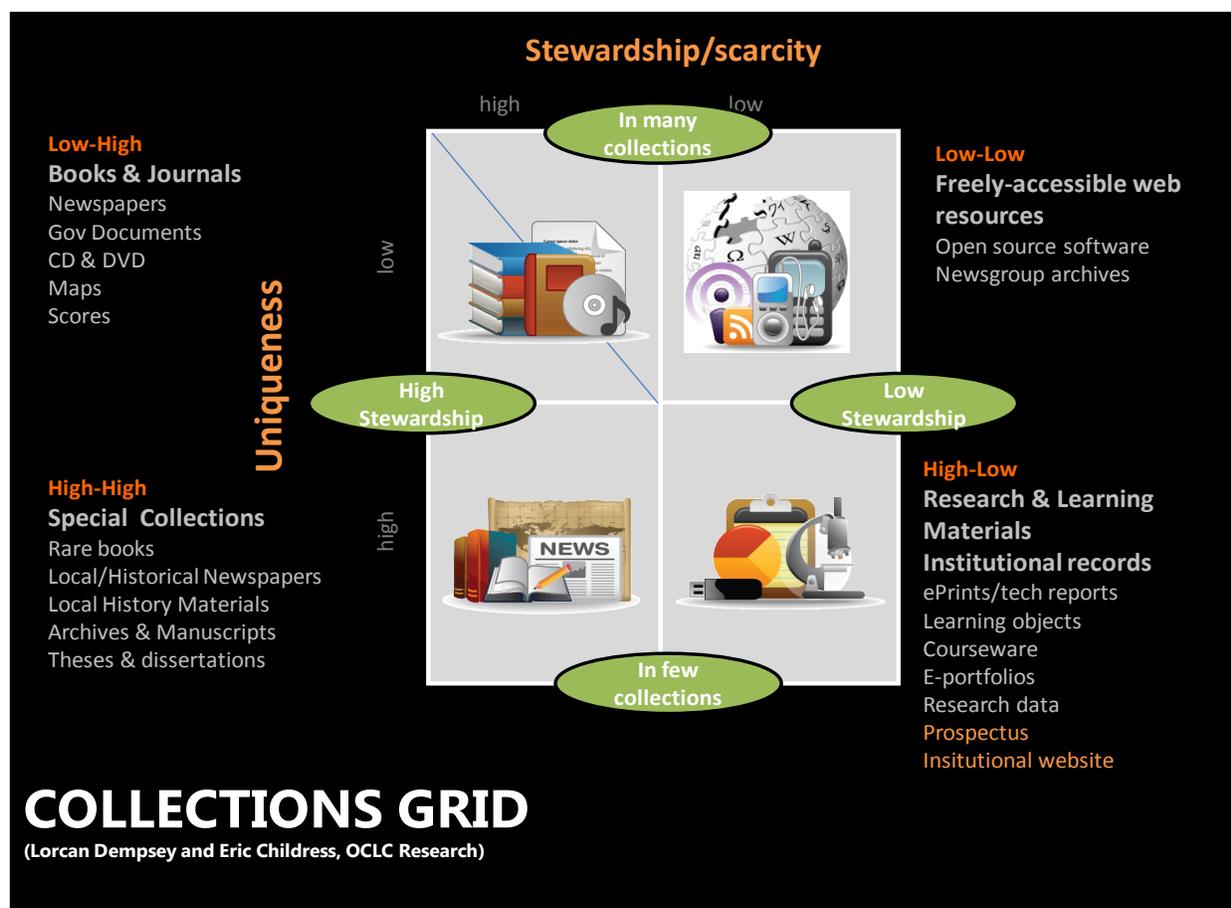

Figure 2. The collections grid

This gives us four quadrants (figure 2).

1. Upper left: **published materials** (books, journals, DVDs, etc.). These are the current core of library collections; they are sustained by an extensive support industry and mature systems and still attract the major part of library staff and resources. They include the broadly available outcomes that form much of the scholarly and cultural record. An important distinction is made between materials which are bought (books, CDs, etc.) and those that are licensed (journals, abstracting & indexing databases, etc.). Until recently that distinction approximately corresponded to print/electronic, but the rise of e-books has changed that. A large part of what is of interest here is commodity material, potentially available through many channels. The boundary between this category and the "special collections" quadrant is somewhat discretionary, although as materials become rarer, more difficult to acquire, or require special processing, they become more "special."

2. Upper right: we put the **open web** here as it can be replicated, indexed, etc., across collections. The boundaries between what is on the open web and what the library provides are being blurred.

3. Bottom left: here is what we know as **special collections and archives**, rare and unique materials which are heavily stewarded. Theses and dissertations, local history materials, and other materials are here. These are attracting more attention because of their reputational value, as new ways are being found to release their value in research and learning, and as they become digitized and potentially aggregated for discoverability.

4. Bottom right: this is a major and growing category. As **research and learning** are carried out in a digital environment, they generate materials (e-prints, research data, learning objects, administrative records, etc.) that need to be managed as institutional assets and disclosed to potential users elsewhere. As these materials become more curated, they assume some of the characteristics of special collections. And indeed, it is interesting to see some new organizational constructs emerge in libraries where responsibility for special collections and new scholarly resources go together.

The value of this approach is that it highlights the different drivers and characteristics in each quadrant. The upper left "published" quadrant is where the traditional "outcomes" of research typically appear. More recently, there is a strong institutional interest in materials "below the line" (special collections, and research and learning materials), raising discoverability and preservation issues in different ways than for materials above the line.

Indeed, together these categories are often part of the "evidence" that form part of the process of scholarly inquiry, whether it is primary materials in the special collections quadrant or research data in the research and learning materials quadrant. These materials present an "inside-out" challenge for the library, and we discuss this further in that section.

Upper Left: Published Materials: Purchased and Licensed

This quadrant has dominated library thinking and organization for many years, as the library acquires and organizes the outcomes of research for local use. It is useful to consider the trajectory of the monograph and journals literature separately. These equate roughly to purchased and licensed materials respectively, although with the emergence of e-books that divide is less clear.

Considering Books

Print continues to be central, but several drivers are altering priorities. These include demands on space, the emergence of a digital corpus through Google Books and HathiTrust, and the cost of managing a resource that releases progressively less value in research and learning. We believe that we are moving to a situation where network-level management of the collective print collection becomes the norm, but it will take some years for service, policy, and infrastructure frameworks to be worked out and evolution will be uneven. At the moment, this trend is manifesting itself in a variety of local or group shared print projects, as well as in several regional and national initiatives. The recognition that system-wide coordination of print materials is necessary as libraries begin to retire collections—to offsite storage or removing them altogether—is gathering. We return to this topic in a separate section below.

> **We believe that we are moving to a situation where network-level management of the collective print collection becomes the norm . . .**

At the same time, the rise of e-books and the shift to licensing models for books is more pronounced. Recent OCLC market research found that e-books are the number one priority of academic library managers in Germany, the Netherlands, the UK, and the US.[9] This is unsurprising, as the long-term trend toward the network delivery of content seems inevitable, though complicated by a variety of thorny issues. These include:

- The potential for scholarly use of e-book content is greater at this stage than what current e-book platforms support.

- The shift from purchase to licensing has raised business model issues, which have an impact on the existing economy and ecology of book use in universities. Think of discussion about fair use, preservation, inter-library lending, and coordinated collection development, as well as discussion about levels of usage and cost.

- Interest in adoption and use varies across discipline and educational stage. For example, one study showed that graduate students and post-doctoral researchers favored electronic formats, while undergraduates preferred print; students in business, medicine, and law had a more positive view of e-books than those in the arts and humanities.[10] Recent Ithaka S+R work has shown that historians, for example, value convenient access to print materials (locally or in an efficient system of delivery) but also value the ability to prospect a large digital corpus, notably Google Books.[11]

Patron-driven models are becoming more interesting to libraries, moving away from the more intentional or curatorial approach.

- Patron-driven models are becoming more interesting to libraries, moving away from the more intentional or curatorial approach.

- This variety of acquisition models raises an interesting new challenge for management and discovery systems, as they need to understand the relationships between print, digitized, and licensed versions of the same content, as well as a variety of fulfillment options.

- Preservation of print books has been largely a local issue, with natural redundancy built in. That is changing in the shared print model, where initiatives are looking at system-wide stewardship. In the licensing model, incentives do not line up in the same way, and preservation becomes an issue.

There is some similarity here with the move to licensing of journals. Notably, the balance between local acquisition and network supply changes and just-in-time models appear. However, it is also the case that there is a very different dynamic at play with journals.

Considering Journals

The heightened awareness around communication of results in science policy, the practical historic sourcing of academic reputation management and validation with publishing organizations outside the academy, and the growth of interest in data alongside new network affordances, have combined to sharpen discussion around the current model of scholarly

publishing, which is an elaborate apparatus of commercial, educational, and not-for-profit elements. This is against a background of the growing volume of research worldwide, changing patterns of research workflow and outputs in a digital environment, and a strong lobby both for open access to the scholarly record and against what are seen as unsustainable publisher pricing models.

Licensed resources now consume the larger part of academic library materials budgets.

Licensed resources now consume the larger part of academic library materials budgets. There is a continued focus on collaborative licensing deals, closer scrutiny of big deals, experiments with just-in-time article delivery models, and support and advocacy for open access and changing scholarly communications models.

At the same time, publishers have broadened their interest to think about research productivity and workflow, research data, researcher identification and profiling, university analytics, and are experimenting with new technologies and business models (including "gold" open access and hybrid approaches). Think for example of the suite of research evaluation and comparison tools provided by Thomson Reuters and Elsevier, which are discussed further below. Or think of the interesting range of experimental services being supported by Macmillan in its Digital Science incubator. These publishers are working to expose their authors, their articles, and research data more effectively on the network. These publishers see journals as one part of a broader ecosystem of services around researcher workflow and research information management in a network environment. Note, for example, Elsevier's recent acquisition of both PURE (for research information management) and Mendeley (to support researcher workflow) in this context.

Libraries advocate for open access on their campuses and provide advice to researchers about options, as well as providing repository support for "green" open access. They advise about and in some cases administer article processing charges. But libraries lack streamlined tools and processes to manage open access materials as part of their collections. There is no reliable way of identifying open access material, there is no consistent practice about identifying use and reuse, and there are no comprehensive aggregations.[12] The transition to hybrid approaches by publishers and the evolving policy context ensure continued change.

Libraries advocate for open access on their campuses and provide advice to researchers about options, as well as providing repository support for "green" open access.

While the journal publishing system is very mixed in terms of publisher types (large and small, for-profit and not-for-profit) large commercial publishers play a major role. Universities have

externalized aspects of academic reputation management and validation to this publishing system—at too high a cost, some would argue—in terms of finance and control. This has led some to argue for a "rebundling" of scholarly publishing into the academy in some way. We return to university/library publishing when talking about the research and learning material quadrant below.

Bottom Left: "Special" Collections

In the bottom left quadrant of the collections grid we cluster a range of materials that are understood to require special curatorial care: rare books, manuscripts and institutional records, maps, audiovisual collections, etc. There is a wish to release special collections more effectively into research and learning activities through better description, digitization, and exhibitions.

These materials are managed separately from the library's general collections, supported by different personnel who rely on separate, sometimes bespoke infrastructure for core operations like resource description, discovery and delivery. While many academic libraries have special collections units, their scope and function differ widely in research universities, where they support a broad stewardship mission, and institutions focused primarily on undergraduate education, where "illustrative" primary source are more likely to be used in instruction. Acquisition of special collections is often supported with special funds, outside of the library materials allocation; gift collections are another major source of growth. As a result, the growth and direction of special collections is somewhat independent of the larger library, which nonetheless provides essential infrastructure for their management.

> **The growth and direction of special collections is somewhat independent of the larger library, which nonetheless provides essential infrastructure for their management.**

Looking ahead, we can predict academic institutions will become more concerned with streamlining the operational management of special collections, so that the costs of continued growth in acquisitions can be moderated and more materials can be made visible. Backlog reduction remains a critical challenge, especially for archival and other special formats for which shared cataloging infrastructure is limited. The sharing of the work burden that is available in traditional cooperative cataloging models (i.e., copy cataloging) is not available for unique or special materials; instead, economies of scale are achieved through simplification and streamlining of descriptive processes and linkages to standard vocabularies.

Two trends are likely to emerge here. First, more deliberate coordination of effort across special collections repositories to share expertise and best practices across specific domains and formats will encourage standardization and leverage collective capacity. Few repositories can provide comprehensive coverage of the literature even in topics for which the library is highly reputed; curatorial and cataloging capacity will need to scale "horizontally" across institutions so that the total value of distributed resources can be realized. Trust and service networks to support this activity are not mature.

Second, the current emphasis on engaging with content users will move upstream, so that interested communities of users participate in the work of resource description. Think for example of the New York Public Library's innovative approach to "gamifying" description of historical maps with its Building Inspector application.[13] Or, in a very different context, think of the UK Public Catalogue Foundation's Art Detective project, which enlists the help of citizens to catalog paintings in public institutions that lack any curatorial or cataloging staff.[14] Even in libraries with significant resources, there is growing acknowledgment that enabling direct and disintermediated access to special collections is an important part of emerging curatorial practice.

With renewed focus on value-based library assessment, there is increased attention to how special collections and archives contribute to research and learning agendas. This has encouraged a stronger focus on how materials are exhibited in the online environment, not just as lists or pictures of "treasures" but as coherent collections of materials that support undergraduate education and advanced research. The special expertise that curators have traditionally directed toward acquisition and management of collections is increasingly turned "outward" to help contextualize and characterize the value of institutional holdings. Consider the example of the University of Illinois's special collections blog *Non Solus*, which highlights particular holdings by embedding them in a larger narrative about specific lines of critical inquiry.[15] Similarly, the University of Texas' *Cultural Compass* blog documents how scholars are using special collections.[16] In this way, the library comes closer to the museum emphasis on exhibition and education.

As a growing volume of digitized special collections is made available outside of general-purpose digital libraries like HathiTrust or Google Books, it will be increasingly important for institutions to leverage other discovery and syndication tools. Harvard's Houghton Library has created a popular Tumblr microblog to highlight digitized special collections in a broader social networking environment.[17] Libraries using digital repository platforms to manage locally digitized content will be more attentive to the importance of Search Engine Optimization (SEO) in maximizing the discoverability of institutional resources.

The very distinctiveness of special collections places important constraints on the development of shared infrastructure or cooperative programs. In a recent survey of special collections repositories in North America, only five percent of libraries reported participating in a formal collaborative collection development program.[18] Yet, nearly all special collections libraries face a shortage of collections space, cataloging backlogs, as well as shared concerns about maximizing visibility of their resources in the digital environment. While it seems unlikely that we will see group-scale infrastructure emerge to support shared management of special collections in the near future, it seems all but certain that increased coordination of collections will be necessary to ensure that operational costs can be better managed. That coordination is likely to rely on metadata aggregations that will need to reflect an increasingly diverse range of formats and descriptive practices, imperfectly accommodated in current record-based library management and discovery systems. New linked data approaches may prove instrumental in supporting richer descriptive practices and enabling better linking of related resources in disparate repositories.

> **Nearly all special collections libraries face a shortage of collections space, cataloging backlogs, as well as shared concerns about maximizing visibility of their resources in the digital environment.**

Bottom Right: Institutional Research, Learning and Profile Materials

This quadrant is increasingly important. We can identify various strands of activity, especially as there is more interest in curating and disclosing more materials from the process of scholarly inquiry, as universities become more aware of the range of digital assets they produce and the management requirements they raise, and as making such assets more discoverable is seen as contributing to university reputation.

Institutional repositories are a routine feature of academic libraries, amid ongoing discussion about purpose and scope, incentives for researchers to deposit, and their role within "green" open access. This is not the place for a full treatment, but a couple of points are worth making. First, while most repositories are home to versions of research papers, scope varies across institutions. For example, some repositories may take a "campus bibliography" approach, including links to publisher splash pages. Some repositories may include other categories of material, institutional records or archival materials, for example. Given the lack of standard methods for designating material types, this may make it difficult for an aggregator of repository content to distinguish scholarly material. Second, there is a close connection between repositories and national education and science policy regimes, so the dynamic of development has been differently influenced in different regimes. For example, where there are national research assessment regimes in place, institutional interest in

repositories may be higher.[19] Shifts in US federal policy with regard to research funding and access to outcomes will have an impact here. This highlights the relationship between the repository and emerging CRIS (current research information system) infrastructure, which is variable depending on institutional configurations.

The institutional repository intersects with a growing university interest in managing information about the research process: research outcomes, grants and income, expertise, and so on. Often, this is led from the institution's Office of Research Management, in support of grants and project administration, tenure and promotion, expertise management and disclosure, and tracking of research outputs. Additionally, research analytics has become of more interest as institutions assess comparative research strengths, collaborations, or compare themselves to peer groups. Bibliometrics may be one strand of this activity. And, as just noted, universities may need to manage information in the context of research assessment or track compliance with funder requirements. This is the context in which we have seen the emergence of the CRIS. Leading CRIS systems are Symplectic Elements and Pure, developed in Europe but now being

The institutional repository intersects with a growing university interest in managing information about the research process: research outcomes, grants and income, expertise, and so on.

deployed elsewhere. Symplectic, originally developed by researchers at Imperial College London, is part of the MacMillan Digital Science portfolio; Atira, the Danish company which created Pure, was acquired by Elsevier in 2012. Pure is aligned with the Elsevier SciVal product in a portfolio labeled "Elsevier Research Intelligence."[20] Thomson Reuters provides Research in View in a category labeled "research analytics."[21] Interestingly, this initiative emerged from a partnership with Ohio State University around their internally built OSU:pro system.[22] Operating in the same space, VIVO provides a community-based approach to managing and disclosing "researcher interests, activities and accomplishments,"[23] and bepress provides repository, communication and research profiling services.[24] In short, there is a growing level of interest around the management, evaluation, and disclosure of research outcomes and expertise, which connects in various ways with internal evaluation and management goals, funding policy and compliance needs, and broader reputation management on the web. We have elaborated a bit here, as this is an important emerging area with which libraries are variably involved.

There has also been growing interest in curating research data and making it available more broadly. There are several motivations for this, including funder mandates and data reuse. There is a very active community of interest here, and an emerging body of best practice (see for example the work of the Digital Curation Centre).[25] Again, the library is potentially a

partner in a multi-stakeholder activity across a campus, and data curation and dissemination has emerged as a major interest for research libraries.[26]

Finally, we mention here library support for faculty and student content creation and publishing.[27] Libraries are being called on to provide support for new modes of scholarly production in a digital environment. Vinopal and McCormick characterize an enterprise array of standard services as follows:

> . . . tools and support teams for activities including high performance computing; geographic information systems; quantitative and qualitative data analysis; data finding and management; the digitization, creation, manipulation, storage, and sharing of media content; repository services; digital preservation; streaming media platforms; digital journal publishing; online collaboration; and intellectual property consultation.[28]

They further note that the library is expected also to support the creation and management of faculty or project-based websites.

Additionally, some libraries recognize a mission-driven role to support open access publishing models as part of the repatriation of the scholarly record to the academy. A recent survey of ARL and other academic libraries noted that "The vast majority of library publishing programs (almost 90%) were launched in order to contribute to change in the scholarly publishing system, supplemented by a variety of other mission-related motivations."[29]

As this area of support matures, some related issues will come to the fore:[30]

Some libraries recognize a mission-driven role to support open access publishing models as part of the repatriation of the scholarly record to the academy.

- A general model to pattern a library's set of responses or service configuration has not yet emerged. So the portfolio of support services will differ from campus to campus. And each library's activity will be differently situated in relationship to other campus services. There is no consistent organizational approach, for example, for the relationship between copyright and other advisory services, library publishing services, university press, research data management, and institutional repository.

- It is sometimes difficult to discern between edge cases and emerging services: for example, are alternative forms of monograph peer review and publication going to emerge as important categories, or will they remain experimental?

- There is a balance between doing extensive custom work for one faculty member or department and the ability to scale services effectively across a campus community.

Vinopal presents an interesting pyramid of services, noting a spectrum from standard enterprise support (e.g., text scanning), to standard research services (e.g., data analysis tools or web exhibits), to enhanced research services (e.g., custom-designed UI), to applied R&D which might be supported by grants.[31]

- Emergent areas live beside established practices. This may lead to a more plural environment or over time to disruption or absorption. Think of the various scenarios that might play out with open access publishing and alt-metrics, for example.

Upper Right: The Open Web

The library collection and the open web might once have been seen as belonging to very different categories. That has changed and will continue to change. Here are some ways in which they overlap, as so much information use has gravitated to the open web.

- There is some systematic selective archiving of web resources, driven by particular subjects or events. Some national libraries do this within their own domain. The work of the Internet Archive has been very important here at a network level, as well as the tools they provide to institutions for web archiving.

- Libraries provide access to web resources in various ways. Metadata may be added to the catalog, for Project Gutenberg books, for example. Links to Google Books may be added. Web resources may be listed in resource guides. As all of these resources get rolled into discovery layers, the distinction between what is in the "collection" and what is not becomes less clear. Discovery and access are decoupled from the local collection. This leads into a general interest in more reliable identification of open access materials. One of the interesting side effects of the MOOC movement, for example, has been to encourage stronger interest in identification and promotion of public domain resources to support courses. It will be interesting to see whether this interest coalesces around new network level discovery aggregations.

> **The distinction between what is in the "collection" and what is not becomes less clear.**

- Libraries want their materials from the bottom left and right quadrants to be found on the web, so they are looking at search engine optimization best practices for being added to Google Scholar,[32] sharing of metadata with aggregators, and so on.

- Libraries are interested in their resources being used appropriately in the "aftermath" of the scholarly process. Licensing has become important in this

context, to indicate appropriate uses and also to reduce the transaction costs of using and reusing materials.

Shifting Emphasis in Institutional Interest: Inside-out and Outside-in

This analysis points us towards a distinction that is quite important in how we think about collections, their management and discovery.

Think of a distinction between outside-in resources, where the library is buying or licensing materials from external providers and making them accessible to a local audience (e.g., books and journals), and inside-out resources which may be unique to an institution (e.g., digitized images, research materials) where the audience is both local and external. Thinking about an external, non-institutional audience, and how to reach it, poses some new questions for the library.

It can be seen from the above discussion that in recent years the institution's unique intellectual assets "below the line" have received more attention, whether these are archives and special collections, or newly generated research and learning materials (e-prints, data, courseware, digital scholarly resources, etc.). As we have already observed, it is interesting seeing the "new" and the "old" come together in this way. Each is a distinctive contribution of the institution; each is the institution's responsibility to preserve to the extent it wishes; each involves use of a metadata and repository apparatus, whether locally created or sourced from the cloud; each involves engagement with learning and research practice in new ways; and each brings to the fore the archival concerns of provenance, authenticity, and context. Each also involves disclosure from the "inside" to an outside world of users; for many of these resources, it is likely that there are more interested users outside the institution than inside it. For this reason, the management of these resources is linked to reputation. And indeed, increasingly, one of the important resources to be managed is researcher expertise and identity. Effective discovery means syndication to search engines, to disciplinary resources, or to other specialist network level resources (e.g., ArchiveGrid, ARTstor).

One of the important resources to be managed is researcher expertise and identity.

Contrast this with the "outside-in" resources. Electronic journal materials are largely managed outside the institution, and are selectively licensed for the institution. However, there is a community wide interest in preservation. They are not necessarily distinctive to the institution. As noted above, the "line" dividing the top half of the quadrant from

the bottom is porous, and one trend to watch is some recalibration of what is "distinctive" to the organization—to include, for example, area studies materials, or other "non-commodity" materials.[33]

The level of attention to "inside-out" resources will become an important differentiator between libraries. Research institutions, specialist libraries, and others with a mission to share their resources with the world will focus attention here. Other institutions more focused on supporting learning and student success may make less of an investment.

As the inside-out focus is an emerging trend, it raises questions that underline some of the main themes of this article.

- **Institutional organization and boundaries**. Some libraries have established explicit digital scholarship support divisions, although as we have already remarked there is not a common pattern. Given the university-wide reach of these materials, they raise some interesting boundary and partnership questions on campus for the library and its relations with other divisions. As the creation, management, manipulation, and disclosure of digital collections of various types have become integral to a wide range of university activities, so have a variety of campus divisions assumed information management roles. Think of campus IT, digital humanities centers, university presses, learning and teaching support, media support, research and data infrastructure in schools and departments, GIS, research management units, and so on. This creates organizational and partnering choices, which will tend to be driven by local personalities and politics, although we can expect to see new patterns emerge.

- **Rightscaling**. What is the balance between institutional activity and subject-based repositories or PubMed instances, for example, in relation to preprints or research data? We have discussed a trend for infrastructure to be unbundled and consolidated in shared platforms, for management, preservation, or discovery. The range of activity discussed here is less mature than others, and we have seen a lot of institutional development. However, there are also signs of shared infrastructure emerging as the network logic becomes apparent. We return to this question below.

- **Discovery**. There is something of a mismatch between discovery requirements for outside-in and inside-out resources. In the former case, the library wants to make known to its users what it has purchased or licensed for them, maybe alongside pointers to other materials. In the latter case, the library often wants to share materials with a broader community, with researchers elsewhere, with professional colleagues, and so on. This places an emphasis on effective disclosure, thinking about search engine optimization, syndication of metadata to network hubs, and so on. The

University of Minnesota has done some interesting work on this question.[34] There is also a desire to have network level discovery venues, which pull together this material. This is done to some extent in Google Scholar, in Worldcat.org, in initiatives such as DPLA and Europeana, and in a range of disciplinary resources such as the ICPSR.[35]

- **Linked data**. Systematically mapping and revealing relationships between entities in archival, published, and digitized collections is of great interest, in both scholarly and cultural heritage contexts. As schema.org markup and other linked data approaches become more common, it will be important to link to authoritative entity backbones to increase the visibility and—of equal importance—the coherence of these collections across the network. Important entities include people, places, and events.

- **Reputation and value shift**. The role of these materials in enhancing the reputation of the institution is an interesting one, and one that is relatively underexplored or quantified. A related issue is the shift in institutional resourcing that will be needed to support an "inside-out" turn in the library. If there is a shift of the type we discuss here, it needs to be justified within the institution, which will require advocacy and persuasion. The case for curation and disclosure of institutional assets is supported in some instances by university mandate or faculty policies (such as required deposit of preprints).

Managing Shared Print

We discuss shared print here because it is an interesting example of how the trends we have discussed will change a historically core library activity. Libraries are beginning to evolve arrangements that will facilitate long-term shared management of the print literature as individual libraries manage down their local capacity. Examples of initiatives here are the WEST Project and the print management activities of HathiTrust. Initially, attention was focused on journal runs, but it is now spreading to monographs as well. Of course, libraries have long worked with print repositories, individually or in shared settings. However, a more systemic perspective is now emerging and we have been using the phrase "collective collection" to evoke this more focused attention on collective development, management, and disclosure of collections across groups of libraries at different levels. We anticipate that a large part of existing print

> We anticipate that a large part of existing print collections, distributed across many libraries, will move into coordinated or shared management within a few years.

collections, distributed across many libraries, will move into coordinated or shared management within a few years.[36] OCLC Research has been working to develop an empirical foundation for this development based on its registry of library holdings, Worldcat.[37]

These shared print initiatives bear an important resemblance with past efforts to develop a shared, library–based infrastructure for the preservation of e-journals in LOCKSS. There is an important difference in that alternative, third-party solutions for e-content preservation (Portico, for example) emerged alongside those organized within the library community. Similarly, the HathiTrust digital preservation archive was developed to meet an emerging need for shared infrastructure to manage the products of large-scale digitization. LOCKSS and HathiTrust represent community-sourced solutions that have enabled academic libraries to externalize stewardship functions that were previously organized locally at a much higher cost. Where print management is concerned, there is little evidence (so far) of third-party interest in developing infrastructure or services to support a new model of library logistics, in which inventory is consolidated and managed as a shared resource. Instead, regional consortia have stepped into the breach to develop policy and infrastructure to support cooperative print management initiatives.

One can consider the emergence of shared print programs as a natural experiment in rightscaling solutions to a shared library problem: how to resize institutional investment in managing library print collections while ensuring their long term preservation for current and future researchers. For academic libraries especially, the opportunity costs associated with traditional, institution-scale inventory management are significant and represent a constraint on institutional innovation.[38] Shared print management schemes represent a cost-effective alternative to institution-scale solutions, redistributing the costs of library stewardship across a broader pool of participants. The scale at which such cooperative regimes are organized will depend upon the economies of scale that can be achieved by pooling inventory, physically or virtually, through shared catalogs and resource sharing schemes, and the economies of scope that can be obtained by diversifying the types of libraries and collections that are included in the shared resource. Shared print arrangements in North America (and elsewhere) largely favor broad regional participation of similar academic library types; they are currently optimized to deliver economies of scale rather than economies of scope. This is understandable, as academic libraries are the primary beneficiaries of the reduced transaction costs, and opportunities for organizational innovation, that cooperative stewardship enables. At the same time, such arrangements risk producing significant dislocation in the library system as a whole, as public libraries and other entities that have relied on tacit agreements with academic libraries are excluded from the decision-making processes that will reshape the emerging collective collection.

Elsewhere, we have described the emergence of regional-scale print management schemes as an important factor in the changing geography of collections.[39] Print inventory and collection management activities that were once organized at institution-scale, are now being reorganized at group scale within regional geographies. A growing number of library consortia are working to concentrate capacity on a regional basis: the Association of Southeastern Research Libraries, the Committee on Institutional Cooperation, the Statewide California Electronic Library Consortium, and the Western Regional Storage Trust, among others, are leading initiatives to develop strategies for managing aggregate print collections as a regional resource.[40] Cooperative infrastructure is being leveraged to support a deliberate recalibration of local and group investment in print management. These efforts will enable participating libraries to redirect attention and investment toward activities that directly support the research and teaching missions of their parent institutions: digitizing and building awareness of distinctive special collections, managing learning objects, supporting scholarly communications.

> **Print inventory and collection management activities that were once organized at institution-scale, are now being reorganized at group scale within regional geographies.**

The shift toward managing legacy print collections as shared assets reflects an important change in library resource management, one that is consistent with our observation that institution-scale stewardship is increasingly focused on materials that are rare or unique. By contrast, materials (whether print or electronic) that are more widely distributed across the library system are more likely to lend themselves to shared stewardship arrangements, whether this is centralized in a single institution or service provider or de-centralized across multiple sites. The growing differentiation in the ways "commodity" and more distinctive library resources are managed has important implications for the organization of libraries considered at both the local and group level.

> **Materials (whether print or electronic) that are more widely distributed across the library system are more likely to lend themselves to shared stewardship arrangements.**

As noted above, the level of attention to institutionally distinctive "inside-out" resources—special collections, locally produced research, teaching and learning objects—will be an important differentiator between research or specialist libraries and those whose primary function is to support "the business of education." Further, we can expect to see those institutions with limited access to the shared infrastructure that supports cost-effective management of "outside-in" collections (shared print facilities, consortial licensing, etc) will have greater difficulty in reconfiguring service profiles to ensure a good fit with the evolving needs of the institutions they serve.

Sourcing and Scaling: The Rightscaling Challenge

As library activities are reconfigured by the network, decisions about rightscaling emerge. By this, we mean that activities may be carried out at various levels (institutional, group, national, etc.) and decisions have to be made about where to invest effort in this context. Shared print provides a good example, as libraries begin to manage down institutional collections, while investing in shared infrastructure at consortial, state, or regional levels.

This is an example of how stewardship responsibility will be more distributed within institutional, collaborative, or third party contexts. We have discussed other examples throughout this article. This means that there needs to be more "conscious coordination" of system-wide responsibilities, where libraries sign up for explicit roles, for example, to collect particular types of material, or where new platforms emerge to concentrate a particular function (for example, see the successive emergence in North America of, among others, OCLC, Ithaka, and HathiTrust). This creates two organizational challenges. The first is to build new trust networks around particular needs, or to extend existing groups; and as discussed below, we do see new groups emerging. The second is that focus is shifted away from institutional resources and services, which may be difficult to justify to local administrations.

A characterization of general trends across the collections grid provides a good context for thinking about this question, and also suggests a framework for rethinking institutional investment in stewardship of the scholarly record. We focus here on the upper left quadrant (published materials), and below the line (special collections and research and learning materials).

We draw a distinction here between the various levels on which collections and related services may be sourced and the scale on which they are made available to an audience. The degree to which various sourcing/scaling arrangements are operationalized is highly dependent on available infrastructure and services.

It is useful to think of three audience scales: local or institution-scale, group-scale (where the intended audience is a consortium, country, or "club" of some sort), and webscale (where the intention is to reach the whole web population). It is also useful to think of three sourcing approaches: local or institutional (where an institution builds or provides a service itself), group (where the approach is collaboratively sourced) and third party (where an external provider is used). For any institution, it is then possible to build a matrix of how they are building out services. In this context, we can say that there is a general trend towards group as libraries try to build scale, and also a growing interest in leveraging webscale providers (think of Google Scholar or of metadata aggregations in DPLA, Europeana, or WorldCat). We show a partially built out matrix for "published" materials below (figure 3).

Upper Left: Published Materials

An array of cooperatively-sourced and third-party solutions is emerging to support institutions "scale up" operations associated with "outside-in," published or commodity collections.

For print materials, shared print management programs like the Western Regional Storage Trust or Maine Shared Collections Strategy are developing policy frameworks and service infrastructure that will enable individual institutions to recalibrate institutional investments in

An array of cooperatively-sourced and third-party solutions is emerging to support institutions "scale up" operations associated with "outside-in," published or commodity collections.

commodity print resources. This is placing previously institution-level decision making in a collaborative group context. Such decisions may also be made with reference to the digitized resources available through HathiTrust. This activity is largely collaboratively sourced.

Source / Audience	Institutional	Group	(Global) Third party
WebScale			Google Scholar; Google Books; WorldCat; Amazon
Group		WEST; OCUL; Orbis Cascade Alliance; HathiTrust	
Local	Course reserves		Locally available purchased and licensed materials.
Audience / Source	Institutional	Group	(Global) Third party

Dempsey et al. for OCLC Research, 2014.

Figure 3. A partially built-out matrix for published materials

Of course this is not an entirely novel phenomenon. In an earlier phase, libraries reduced their local management of the journal literature in favor of a licensing model. This in turn created shared "group-scale" approaches to negotiation, preservation, and advocacy around new scholarly communication approaches. It also resulted in the emergence of a variety of aggregator and other supply side services, packaging materials for library or researcher needs. This trend has continued, and libraries are increasingly balancing licensing decisions against demand-driven models and just-in-case models, serviced by various providers.[41] In this way we can see the emergence of group-scale licensing and negotiation with third-party sourced materials.

Commodification here means that it is not difficult to imagine career- or some teaching-focused institutions acquiring much of its library service through a small number of providers (EBSCO, for example) and making limited investment in local customization of these resources. In other words, they unbundle a large part of the library service to a third party provider.

Discovery of these collections provides an interesting "rightscaling" question also. We have seen a shift of emphasis from focused institutional attention to the catalog, to, in some cases, various group or consortial discovery systems (e.g., OhioLink), to "web-scale" discovery services that aggregate data from various sources, to broader network hubs like Google Scholar. Increasingly, the library will have to think about how it is represented at all of these scales.

Finally, we have seen a strengthening of group-scale services for libraries that are in consortial arrangements. Here, a range of collection, management, and discovery functions may have partially moved to the consortial level. Consider OCUL (Ontario Council of University Libraries), for example, or OhioLINK in this context. There are many organizational and service constructs here, but participating libraries balance local and consortial activity across a range of areas. Some of these groups

We have seen a strengthening of group-scale services for libraries that are in consortial arrangements.

are looking at shared print needs. Of course, these organizations are also looking at repository and other digital infrastructure needs. As noted above, these organizations provide a framework within a more data-driven network.

Below the Line: Special Collections, Research and Learning Materials

The infrastructure needed to support the rightscaling of operations for special collections and research/learning resources is fragmented. Where format-specific tools and services exist,

they are imperfectly integrated with the infrastructure that supports management of other core business functions. We anticipate that a complex environment will continue given the diversity of activity, but that shared platforms will emerge in important areas.

Consider the example of resource description and discovery services for archival collections, which can be commercially or cooperatively sourced. Purpose-built solutions like CALM, Archeon, or ArchivesSpace sit alongside the local library management systems and streamline institution-scale work; efficiency gains are felt locally, but not as a shared benefit across libraries. There is some group national infrastructure to provide discovery and management support, for example, Archives Hub[42] in the UK or Calames[43] in France. ArchiveGrid, a Web-based discovery application, provides a system-wide view of archival collections in hundreds of repositories; it serves as a syndication hub for archives, operating at some remove from related back-office business operations like reference or fulfillment.

There is a variety of institution-scale repository approaches, sourced in a variety of ways (locally developed, collaboratively-sourced through various open source projects, and sourced from third party providers). While distinctions are not routinized, one might note some overlapping specializations. Several repositories are in common usage for digitized special collections, images, and cultural heritage collections. These may have support for media viewing, digital exhibits, and so on. Institutional repository software is often based on a particular open source offering. Greater support for CRIS-like functionality is one direction. There may be another type of support for research data, for video, or for specialized scholarly products, and a platform for open access publishing.

This is a very heterogeneous environment, although one can note some trends. One is the interest around the Fedora/Hydra/Blacklight stack for those that favor open source.[44] Another is the availability of hosted solutions from various providers (for example, CONTENTdm, ePrints, Duraspace). Another is the emergence of brokered frameworks for archiving services. For example, Jisc in the UK has developed a "data archiving framework," which reduces the transaction costs of finding and negotiating for reliable data archiving capacity.[45] Similarly, DuraSpace provides DuraCloud, a managed service for archiving data with various backend suppliers.[46]

The Jisc case mentioned above is an example of emerging group infrastructure, as a national provider works to build capacity for the system as a whole. DANS in The Netherlands provides national-level data archiving services.[47] The Australian National Data Service is a collaborative response to data needs.[48] In the US, we have seen the nascent Academic Preservation Trust (APTrust) and Digital Preservation Network (DPN) emerge as shared venues for coordinated preservation. APTrust, a consortium of leading US research libraries, is advancing work on a shared preservation repository in which research materials from many

universities will be aggregated. In parallel, DPN is developing a federation of independently governed repositories. While these initiatives cannot fill the gap in cooperative or commercially sourced solutions for acquiring, managing, or providing access to the full range of resources "below the line" in the collections grid, they aim to considerably improve rightscaling options for universities.

As a growing share of the scholarly publication record is made available through open access (OA) channels, libraries are also recognizing the need for common infrastructure to manage OA content. Research libraries in the US and UK are exploring different approaches to coordinating national open access repositories to reduce the burden on individual institutions, in the ARL SHared Access Research Ecosystem (SHARE) and RLUK Open Mirror projects.[49] There is a range of other frameworks. The Narcis service in the Netherlands is interesting, providing a national service, which makes information about publications, data sets, research, people, organizations, and enhanced publications available.[50] There is some coordination of national and other repository frameworks through COAR (Confederation of Open Access Repositories).

> **As a growing share of the scholarly publication record is made available through open access (OA) channels, libraries are also recognizing the need for common infrastructure to manage OA content.**

Discovery is another example. Group-scale discovery services are still limited and tend to be organized by discipline: think of RePEc (economics), arXiv (physics) or SSRN (social sciences). Social discovery platforms (Academia.edu, Mendeley, ResearchGate, etc.) enable researchers to share papers within and across disciplinary communities.

There is no single gravitational center around which group-scale operations for digital repositories are organized. This may change for research materials as mandates for public access to publicly-funded research are implemented. At the same time, prominent federations of digitized cultural heritage content (e.g., Digital Public Library of America, Europeana) are exploring sustainability options. WorldCat provides access to collective library collections, including those digital materials that libraries choose to share.

Finally, although we have considered the right and left quadrants of the collections grid together here, there is an important difference. There will be more commercial and cooperative options on the right side of the diagram. There is stronger commercial interest and greater cooperative capacity in the broad educational space. Libraries share the left quadrant with the cultural heritage sector, which in general is less well resourced.

This diversity is a natural situation early in the lifecycle of an industry or community. There is local innovation, and early adopters gain valuable experience. Some institutions will continue to build local infrastructure. However, it is also clear that there will be increasing consolidation around shared platforms, as institutions seek to gain the efficiencies of scale and the gravitational pull of a consolidated presence on the network. The current environment is sub-optimal from the perspectives of both institutional efficiency and impact/discovery. We anticipate major policy, national and consortial, or group attention to these questions in coming years.

> **There will be increasing consolidation around shared platforms, as institutions seek to gain the efficiencies of scale and the gravitational pull of a consolidated presence on the network.**

Some Concluding Remarks

The general theme of this article has been that the network is reconfiguring not only individual academic libraries but the whole library system. This is because reduced transaction costs facilitate the unbundling of functions and their consolidation in network platforms and specialist providers. There are two key ways in which the network is reconfiguring the library.

1. On the supply side, libraries are moving "above the institution" and becoming increasingly embedded in networks of collaboration, cooperation, and consolidation that are fundamentally changing the ways in which collections and related infrastructure are developed, managed, and made accessible. This means that they are drawn into discussions about scaling (Where should collections management be operationalized, locally or within a group?) and sourcing (Should we build or buy? Should we collaborate with other libraries or source from a third party?).

2. On the demand side, faculty and students are operating in a network environment that is now rich in available resources. The transaction costs of discovering and accessing resources have been significantly reduced. As the centrality of the local library collection is reduced, libraries are more deeply engaging with broader research and learning workflows and touching more points along the lifecycle of creation, curation, and access. Management and effective exposure of institutional research and learning materials becomes more important. The library becomes more interested in supporting creation alongside curation and consumption.

In this context, we anticipate a reorganization of academic and especially research library collections, driven by the following changes:

- Ongoing policy attention to the outputs of publicly-funded research, the pressure to contain costs and to redistribute attention to greater engagement with research and learning, and the influence of the digital network will continue to drive change.

- Discovery will continue to be progressively decoupled from the local collection as "facilitated access" emerges as core a service, which is not necessarily anchored in a local collection.

Universities, publishers, and other service providers are all diversifying research and learning support services that support multiple disciplinary communities.

- The scholarly record is diversifying to include both the traditional outcomes of research and the products of enquiry (primary materials, data, methods, preprints, etc.), as well as derivative, repurposed, and aggregate works.

- This is driving an interest in an ecosystem of services organized around research workflows, discovery, communication, and assessment. Universities, publishers, and other service providers are all diversifying research and learning support services that support multiple disciplinary communities.

- As the library becomes more engaged in research and learning workflows, in supporting the creation, curation and disclosure of institutional research and learning materials, it needs to rebalance investment in "commodity" materials.

- This drives an interest in rightscaling the investment in print, shifting it from local to shared environments, and licensed electronic resources, moving to consortial and demand-driven licensing models.

- To increase operational efficiencies, library workflows will need to be more intelligent and data-aware, using demand-side usage data to trigger acquisitions, collection balancing between institutions, triage for digitization, consolidation in shared print environments, transfer or withdrawal decisions, and so on.

- An inside-out orientation will become more important as universities focus attention on distinctive institutional assets and libraries direct increased curatorial attention toward special collections, new scholarly products, research preprints, and pedagogical resources.

Ultimately, the degree to which these broad environmental changes will affect the organization of academic libraries will depend upon the availability of appropriate collaborative infrastructure (for "above the institution" or "network level" discovery, preservation, collection evaluation, and the like) to support the "conscious coordination" of core operations. Building shared services at scale is necessary and a challenge. Above all, this reorganization will require that academic libraries focus on the distinctive roles they have and deliberately direct attention and resources toward services that deliver value to the institutions and academic constituencies they serve.

Building shared services at scale is necessary and a challenge.

Notes

1. These opening remarks draw on: Dempsey, Lorcan. 2012. "Libraries and the Informational Future: Some Notes." *Information Services and Use*. 32 (3) 203-214.
2. In the US, OCLC has consolidated activity at the national level. In Europe a variety of national and other organizations were formed at the same time, many of which still exist in some form: Dempsey, Lorcan. 1992. *Library Bibliographic Networks in Europe: A LIBER Directory*. The Hague: NBLC.
3. Castells, Manuel. 2010. *The Rise of the Network Society*. Malden, MA: Wiley-Blackwell. "1-597 passim"
4. Moretti, Franco. 2013. *Distant Reading*. London: Verso.
5. Palmer, Carole L., Lauren C. Teffeau, and Carrie M. Pirmann. 2009. *Scholarly Information Practices in the Online Environment: Themes from the Literature and Implications for Library Service Development*. Dublin, Ohio: OCLC Research.
    http://www.oclc.org/content/dam/research/publications/library/2009/2009-02.pdf
6. This phrase is from: Hagel III, John and John Seely Brown. 1999. *Institutional Innovation: Creating Smarter Organizations to Scale Learning*. New York, NY: Deloitte University Press.
    http://www.scribd.com/doc/129958072/Institutional-Innovation-Creating-Smarter-Organizations-to-Scale-Learning.
7. This section is based on: Lavoie, Brian, Eric Childress, Ricky Erway, Ixchel Faniel, Constance Malpas, Jennifer Schaffner, and Titia van der Werf. 2014. The Evolving Scholarly Record. Dublin, Ohio: OCLC Research. http://www.oclc.org/research/publications/library/2014/oclcresearch-evolving-scholarly-record-2014.pdf.
8. There is some similarity between this analysis and Rick Anderson's recent discussion of collections, although we discuss a fuller range of materials. Rick Anderson, 2013. *Can't Buy Us Love: The Declining Importance of Library Books and the Rising Importance of Special Collections*. New York, New York: Ithaka S+R.
    http://www.sr.ithaka.org/sites/default/files/files/SR_BriefingPaper_Anderson.pdf&nid=614
9. OCLC. 2013. "Libraries in the U.S.: A Snapshot of Priorities & Perspectives." Accessed 4 May 2014. https://oclc.org/reports/us-libraries.en.html.

    OCLC. 2013. "Libraries in the UK: A Snapshot of Priorities & Perspectives." Accessed 4 May 2014. https://oclc.org/reports/uk-libraries.en.html.

    OCLC. 2013. "Libraries in Germany: A Snapshot of Priorities & Perspectives." Accessed 4 May 2014. https://oclc.org/reports/de-libraries.en.html.
10. University of California Libraries. 2011. *UC Libraries Academic e-Book Usage Survey: Springer e-Book Pilot Project*.
    http://www.cdlib.org/services/uxdesign/docs/2011/academic_ebook_usage_survey.pdf
11. Schonfeld, Roger C., Jennifer Rutner. 2012. *Supporting the Changing Research Practices of Historians*. New York, New York. http://www.sr.ithaka.org/sites/default/files/reports/supporting-the-changing-research-practices-of-historians.pdf.
12. See the interesting survey results presented here: Bulock, Chris and Nathan Hosburgh. 2014. "OA in the Library Collection: The Challenge Of Identifying And Managing Open Access Resources."

Presented at the NASIG 29th Annual Conference, Ft. Worth Texas, 2 May, 2014.
http://www.slideshare.net/NASIG/oa-in-library-collection.

13. http://buildinginspector.nypl.org/
14. http://thepcf.org.uk/artdetective/
15. http://nonsolusblog.wordpress.com/
16. http://blogs.utexas.edu/culturalcompass
17. http://houghtonlib.tumblr.com/
18. Dooley, Jackie M., and Katherine Luce. 2010. T*aking Our Pulse the OCLC Research Survey of Special Collections and Archives*. Dublin, Ohio: OCLC Research.
http://www.oclc.org/research/publications/library/2010/2010-11.pdf.
19. MacColl, John. 2010. *Research Assessment and the Role of the Library: A Companion Report to: A Comparative Review of Research Assessment Regimes in Five Countries and the Role of Libraries in the Research Assessment* Process. Dublin, Ohio: OCLC Research.
http://www.oclc.org/research/publications/library/2010/2010-01.pdf.
20. Elsevier. "Elsevier Research Intelligence." Accessed 2 May 2014. http://www.elsevier.com/online-tools/research-intelligence.
21. Thomson Reuters. 2014. "Research Analytics." Enterprise Systems. Accessed 2 May 2014.
http://researchanalytics.thomsonreuters.com/solutions/enterprisesystems/.
22. The Ohio State University. "OSU:pro Becomes Research in View." In: Office of the Chief Information Officer. Posted by Rake.9@osu.edu. 1 December 2011.
https://ocio.osu.edu/blog/community/2011/12/01/osupro-becomes-research-in-view/.
23. VIVO. 2014. "About." Accessed 2 May 2014. http://www.vivoweb.org/about.
24. bepress.2014. "About bepress." Accessed 4 May 2014.
http://www.bepress.com/aboutbepress.html.
25. DCC (Digital Curation Centre). Accessed 2 May 2014. http://www.dcc.ac.uk/.
26. Erway, Ricky, 2013. "Starting the Conversation: University-wide Research Data Management Policy." *Educause Review*. December 2013. http://www.educause.edu/ero/article/starting-conversation-university-wide-research-data-management-policy.
27. Kristi Palmer. 2010. "New Models of Scholarship & Publishing." In *ACRL Scholarly Communication Toolkit*. 15 October 2010. Accessed 10 October 2014.
https://web.archive.org/web/20140803110204/http://www.scholcomm.acrl.ala.org/node/7.
28. Vinopal, Jennifer, and Monica McCormick. 2013. "Supporting Digital Scholarship in Research Libraries: Scalability and Sustainability." *Journal of Library Administration*. 53 (1): 27–42.
doi:10.1080/01930826.2013.756689.
29. Mullins, James L., Catherine Murray Rust, Joyce L. Ogburn, Raym Crow, October Ivins, Allyson Mower, Daureen Nesdill, Mark Newton, Julie Speer, and Charles Watkinson. 2012. *Library Publishing Services: Strategies for Success: Final Research Report*. Washington, DC: SPARC.
http://wp.sparc.arl.org/lps/.
30. These are based on: Dempsey, Lorcan. "Three Challenges: Engaging, Rightscaling and Innovating." *Lorcan Dempsey's Weblog* (blog). Posted 5 July 2013.
http://orweblog.oclc.org/archives/002213.html.
31. Vinopal, Jennifer. 2014 "SERVICES! For All Areas of Knowledge!" *Jennifer Vinopal* (blog). Posted 9 January 2014. http://vinopal.org/2014/01/09/services-for-all-areas-of-knowledge/.
32. Arlitsch, Kenning and Patrick S. O'Brien. 2013. *Improving the Visibility and Use of Digital Repositories through SEO*. Chicago: ALA Techsource.
33. While we have used "commodity" in relation to the collections grid, our use of the term here is also influenced by Rick Anderson's use: Rick Anderson. 2013. *Can't Buy Us Love: The Declining Importance of Library Books and the Rising Importance of Special Collections*. New York, New York: Ithaka S+R.
http://www.sr.ithaka.org/sites/default/files/files/SR_BriefingPaper_Anderson.pdf&nid=614.

34. Fransen, Janet, Deborah Ultan Boudewyns, Cody Hanson, Heather Hessel, Lara Friedman-Shedlov, and Stacie Traill. 2011. *Discoverability: Investigating the Academic Library's Changing Role in Connecting People to Resources*. Retrieved from the University of Minnesota Digital Conservancy. http://purl.umn.edu/116586.
35. Inter-university Consortium for Political and Social Research. Accessed 4 May 2014. http://www.icpsr.umich.edu.
36. Dempsey, Lorcan. 2013. *The Emergence of the Collective Collection Analyzing Aggregate Print Library Holdings*. Dublin, Ohio: OCLC Research. http://www.oclc.org/content/dam/research/publications/library/2013/2013-09intro.pdf.
37. Dempsey, Lorcan, Brian F. Lavoie, Constance Malpas, Lynn Silipigni Connaway, Roger C. Schonfeld, JD Shipengrover, and Günter Waibel. 2013. *Understanding the Collective Collection: Towards A System-wide Perspective on Library Print Collections*. Dublin, Ohio: OCLC Research. http://oclc.org/content/dam/research/publications/library/2013/2013-09.pdf
38. On the economic costs of print management, see Courant and Nelson. 2010. *The Idea of Order: Transforming Research Collections for 21st Century Scholarship*. Washington, D.C.: Council on Library and Information Resources.
39. Lavoie, Brian F., Constance Malpas and J.D. Shipengrover. 2012. *Print management at "Mega-scale": a Regional Perspective on Print Book Collections in North America*. Dublin, Ohio: OCLC Research. http://www.oclc.org/research/publications/library/2012/2012-05.pdf.
40. CRL maintains an inventory of shared print initiatives in the Print Archives Preservation Registry: http://papr.crl.edu/. Accessed 4 May 2014.
41. Goedeken, Edward A., and Karen Lawson. 2014 (preprint). "The Past, Present, and Future of Demand Driven Acquisitions in Academic Libraries." *College & Research Libraries*. Manuscript#: crl14-579. http://crl.acrl.org/content/early/2014/03/31/crl14-579.short.
42. Archives Hub. 2014. Accessed 6 May. http://archiveshub.ac.uk/.
43. Abes. "The Calames Catalog." Accessed 6 May 2014. http://en.abes.fr/Calames/The-Calames-catalogue.
44. See, for example, the discussion of the design of digital services at Stanford: Cramer, Tom, and Katherine Kott. "Designing and Implementing Second Generation Digital Preservation Services: A Scalable Model for the Stanford Digital Repository." *D-Lib Magazine*. 16(9/10). doi:10.1045/september2010-cramer.
45. Jisc. "Data Archiving Framework." Accessed 6 May 2014. https://www.ja.net/products-services/janet-cloud-services/data-archiving-framework.
46. DuraCloud. Accessed 5 May 2014. http://www.duracloud.org/.
47. DANS (Data Archiving and Networked Services). Accessed 5 May 2014. http://www.dans.knaw.nl/en/content/about-dans.
48. ANDS (Australian National Data Service). Accessed 4 May 2014. http://www.ands.org.au/about-ands.html.
49. The full ARL SHARE proposal is available here: AAU, APLU, and ARL. 2013. SHared Access Research Ecosystem (SHARE): Development Draft Proposal. 7 June. http://www.arl.org/storage/documents/publications/share-proposal-07june13.pdf.
A brief description of the RLUK Open Mirror co-design project with Jisc is here: Jacobs, Neil. 2013. "What Could You do with an Open Mirror?" *Jisc* (blog). Posted 22 October. http://www.jisc.ac.uk/blog/what-could-you-do-with-an-open-mirror-22-oct-2013
50. Narcis. Accessed 6 May 2014. http://www.narcis.nl/.

For more information about our work, please visit:
www.oclc.org/research

 OCLC® Research