Lost in the Cloud

Research Library Collections and Community in the Digital Age

Dan Hazen

Digital technologies, renewed attention to the purposes of higher education, and changing models for scholarship and learning challenge our historic understandings of research libraries and their collections. Common assumptions and goals are giving way to diverse local agendas, many of which also reflect increasingly limited budgets. Cooperative ventures are taking new forms as well, with straitened resources again the rule. Our adaptation to this uncertain environment requires research libraries to reconsider the elements that are now necessary for success.

Research libraries come in many sizes, offer a variety of services, and support institutions with diverse programs and styles.[1] Despite their differences, these libraries until recently regarded collections as their primary focus. This shared sense of purpose, however, is now in question. The staggering growth and variety of information resources challenge our collective mandate to track, organize, and preserve the full records of scholarship and human expression. Ongoing shifts in the practice of research have made even the largest collections inadequate to many needs. Digital technologies are transforming the nature of information and with it the research questions we ask, the ways we seek answers, and how we communicate results. Academic libraries also support instruction, a high-stakes activity that today requires new types of understanding and engagement. All library operations are constrained by tightening budgets, marketplace economics, and restrictions on intellectual property. Individual research libraries are grappling with this unwieldy mix in disparate ways and often in isolation. The consequences may weaken them all.

This essay reviews the overlapping transformations in technology, information and its availability, scholarship, and instruction that define the research library environment. The information marketplace injects another dimension of complexity. While institutional responses make sense at the local level, they together comprise a cacophony of divergent programs and goals. Active acknowledgement of a few broad considerations may revitalize a sense of common purpose and a capacity for collective success.

A Community in Flux: Digital Fault Lines and the Emergent Research Library

Predigital research libraries seem almost absurdly simple today. Their main role was to acquire the largest possible array of locally relevant books and journals and

Dan Hazen (dchazen@fas.harvard.edu) is Associate Librarian for Collection Development, Harvard University, Cambridge, Massachusetts.

Submitted January 31, 2011; tentatively accepted March 15, 2011, pending modest revision; revision submitted March 25, 2011, and accepted for publication March 27, 2011.

then interpret them to users on-site. Libraries, like universities, were bounded and physical. Postsecondary instruction and research were tightly framed by each field's knowledge base and methodologies. The academic enterprise, particularly in the humanities and social sciences, centered on canonical sources and core texts; nonprint resources supported more specialized domains. Libraries' hardcopy holdings, plus complementary university collections of paintings and plant specimens, cultural artifacts and animal bones, minerals and musical instruments, and so on, reflected the full records of human creativity and natural diversity. Scholarship depended on direct access to these materials.

Research libraries achieved status in this environment by acquiring more than their peers or by building niche collections of particular depth. The measures of performance were clear and rankings made intuitive sense. Collections cooperation was largely limited to esoteric fields or to circumstances of unusual geographic proximity. The library community somewhat fuzzily aspired to collect comprehensive holdings of relevant materials. While the dimensions of "relevance" expanded over time, libraries' collections-centered conceptual universe was largely static.

Several interlocking shifts have brought complexity and uncertainty to this once-placid scenario. Electronic technologies have made information abundant rather than scarce and ubiquitous rather than bounded by its physical containers. Research and scholarly communication have evolved accordingly. The mantra of accountability has moved universities to focus on pedagogical performance. Libraries likewise perceive new opportunities as they also revisit old practices that no longer make sense. The following sections address these three dimensions of change. The focus then shifts to the economic considerations that are affecting collections and services now that profit-seeking pervades large swaths of the information and entertainment landscapes.

Research and Scholarly Communication in the Digital Environment

Digital technologies, in reshaping the information landscape, also have altered the relationship between recorded knowledge and the activities of research and teaching. The Internet, broadly considered, deploys technology in ways that encourage open participation and easy expression. These potentially liberating features, however, mesh poorly with some of the scholarly community's more buttoned-down needs.

Most scholarship, even today, builds from discipline-specific processes that promote documented, reproducible results. Research findings are validated through the judgment of peers and then made broadly available to fuel further inquiry. Kuhn perceives a larger pattern of punctuated equilibria in which periods of model-driven "normal science" are interrupted as anomalous findings provoke disruptive paradigm shifts.[2] New theories and explanatory syntheses then allow the cycle to begin again. Rigorous debate is part and parcel of this dialectic through which self-defined peer communities establish and enforce common evaluation criteria and research agendas.

Several characteristics of the Internet are at odds with so orderly a model for scientific discovery and scholarly communication. Low barriers to participation invite broad inclusiveness and flamboyant individualism. Anonymity is simple: the authenticity of identities, the reliability of sources, authorial accountability, and the credibility of particular assertions therefore are all up for grabs. The Internet is also vast, making the mechanisms by which specific materials can be found a matter of critical importance. Today's approaches to discovery rely on either purposeful prior combinations of separate resources or standardized protocols that assemble dispersed information virtually. Search algorithms, which are often opaque, figure large in determining what we can know and how we can know it. And then the Internet is a flat and epistemologically uncertain terrain. Most search results decontextualize individual listings, obscuring the relationships between them. A snippet from a book, an article abstract, a newspaper account, an archival excerpt, a blog posting—all carry the same valence, conveying a sensation of relentless parity.

The open web—the Internet, the cloud—therefore is in some crucial aspects flawed as a tool for scholarly communication, even as it speeds exchanges and widens participation. While different disciplines value timeliness, rhetorical polish, and reporting protocols in their own particular ways, all insist on documented and reproducible findings, and all expect full recognition of authors and sources. Specific communities of practice are thus exploring approaches that take advantage both of new possibilities (for instance, free-wheeling discussion lists or blogs) and more tightly managed venues for structured debate.

Of course the Internet is much more than a vehicle for scholarly exchange. Its extravagant openness and its utility for communication also make it ever more important as a direct source of information, opinion, creative expression, and data. Vast sweeps of digital primary sources are now available on a scale hitherto beyond imagination. Analog resources, the stuff of traditional library collections, remain crucial as well. In most fields, scholarship requires both types of content.

Educational Accountability and User-Centric Instruction

The Internet's role in scholarly communication is still taking shape—by and large one field at a time. Its utility for teaching and learning may be easier to exploit. The larger context is one in which accountability has become

all-important. Higher education is thought to perform poorly. Colleges and universities are blamed for indifferent educational outcomes, low graduation and placement rates, soaring costs, and misplaced priorities. Many proposals for improvement begin with student learning, variously calling for personalized programs of study, high-end technologies, and collaborative pedagogies.

Contemporary research on student expectations and lifestyles indicates that speed, seamlessness, and accessibility are taken for granted.[3] For many, life plays out through social media as well as private interactions. Virtual worlds, mobile devices and apps, and the omnipresent Internet itself reinforce a context of constant stimulation and instant gratification.

Teachers and librarians find themselves on the front lines as calls for measurable results converge with changing student behaviors. Both groups have always sought to imbue students with the facts and methodologies of domain-specific knowledge, hone their skills of critical thinking and clear expression, and strengthen the abilities that we label as information and media literacy. Specific expectations—whether everyone today should be able to create multimedia presentations, for example—as well as matters of teaching techniques and measurable outcomes, pose ongoing challenges. Digital capabilities are central to our answers as well as our questions. For instance, new technologies can narrow the gap between those who maintain that learning requires mastery of particular, predetermined sets of facts and modes of analysis, and those who link educational success to the active participation of students who are approached in terms of their own styles and tastes.

The drive toward educational accountability, student-centered learning, and instrumental understandings of results, however, raises doubts as well as possibilities. For many, education is about much more than economic vitality or returns on investment. The impact of instruction in the liberal arts, in particular, is difficult to measure—and perhaps is not meant for the sorts of assessment currently in vogue. What today is described as "long-form reading," extended texts that rely on linear arguments and a cumulative flow, meshes poorly with a participatory culture of mash-ups, social networks, and mobile apps. The media and formats through which knowledge is shared, as well as learning outcomes and pedagogical methods, are in question. Libraries fit into these scenarios in different and sometimes uncertain ways.

Libraries in the Cloud

Research libraries are central to the interplay between information and scholarship. Analog library collections, through purposeful selection, sought to locally recapitulate the intellectual domains claimed by particular fields of study. Although the bounds of these ideal collections expanded over time, reflecting increasingly broad disciplinary paradigms for scholarly work, their underlying purpose remained the same.

The Internet and other digital technologies allow a different perspective. Digital resources need not, and in some cases cannot, be held locally. Content is amorphous, inclusive in its formats and modalities of use, and close at our fingertips. With such vast amounts of information available without intervention or intermediary, a collections logic of enforced parsimony and conscious selectivity can feel anachronistic and even perverse. While most disciplines continue to honor their canonical cores, direct access to everything we can assemble seems, on its face, an absolute good.

The library's mediating role is not limited to its once-clear mandates concerning collections and content. Libraries have always been in the forefront of organizing knowledge. Tools like thesauri, taxonomies, classification systems, and cataloging rules can in theory be applied regardless of an object's format, and thus should work as well for information in the cloud as for printed resources. External applications of taxonomies and metadata, however, are slow and expensive. Automated processing, topical ontologies and folksonomies, and the Semantic Web have not yet fulfilled their potential. In the meantime, search engines have eclipsed bibliographies, catalogs, and other scholarly aids as launching pads for research and learning. "Search" privileges the particular with its focus on keywords and specific data, promoting a rigorous granularity that encourages fact-finding; it also encourages data mining and some kinds of meta-analysis. The library's potential role in more sensitively organizing web-based knowledge is neither visible nor appreciated.

Research libraries are under pressure to adapt their traditional functions—identifying, acquiring, organizing, interpreting, and preserving relevant resources—to the web's modes of presentation and delivery. The scale and ubiquity of cloud-based information and services can make this agenda seem simplistic and naive. The network effects of massive amounts of distributed information likewise challenge our notions of organizational roles, institutional structures, and operational efficiency. The community's preservation mandate particularly may be slipping from view, both as the function scales beyond the reach of individual institutions and as collecting priorities leave behind the "long tail" materials that may be most at risk.[4] Finally, even ongoing service norms may warrant new thinking. Research libraries' insistence on providing access to electronic resources that are invisible as well as seamless, for instance, may obscure their role in the process. Some forms of patron-driven functionality can have the same effect. Fluid transactions are possible as never before, but transactional transparency may be a misplaced goal.

Research libraries, separately and as a group, face existential questions of whether and how they fit in the emergent Internet environment. Some libraries have chosen to focus on new information behaviors, staking out areas in which revamped services can provide added value. Others emphasize the continuing weight of analog resources, even as they build digital capacities. Affluent institutions seek to leverage their privileged position through collections and services that will continue to attract the best students and scholars. The broad community's presumed needs for preservation and large-scale coherence are less certainly in the mix. Economic constraints limit the options available either to individual institutions or to the community as a whole.

Scholarship, Libraries, and the Economics of Information

Most colleges and universities face very tight budgets. Research libraries are further affected by the problematic economics of knowledge. Even as the Internet offers vast amounts of unmediated and freely available content, the challenges of commodified information continue to grow.

Price Pressures

Robust scholarship presumes an unobstructed flow of ideas and information. Faculty members (and others) transmute their research findings into books and articles, and this expanding record of scholarship supports new rounds of knowledge-making. Publishers contribute to the process by evaluating, editing, packaging, and distributing the scholarly record. Libraries provide users with access to these materials, typically ensuring that reliable versions remain available over time. While each actor may attach a different value to the elements in this system, it distills a long history of specialization and refinement.

This model is under pressure as aggressive pricing practices and restrictive licenses are applied to digital products; the information economy has engendered a new economics of information. Price increases have consistently and significantly outpaced general inflation, and content providers have imposed severe restrictions on sharing and use. Sufficient funding would, in principle, still allow libraries to acquire or provide access to even the most expensive and closely held materials. Inadequate budgets force less generous results.

The controversy around price distortions in the e-journals marketplace is both complex and, by now, familiar.[5] Major elements include the linkages between publications and scholarly reward; the demographics of the research sphere; provisions for intellectual property; disputed claims of value, prices, and profits; and the propriety of particular content strategies and marketing tactics. A brief rehearsal will set the stage for a few observations on the community's response.

Scholars pursue knowledge and truth, and also promotion, tenure, and prestige. Readily available, peer-reviewed articles and books from high-profile outlets are especially prized. As more researchers enter the fray, new academic specializations and additional publications inexorably follow. Our system of scholarly communication has evolved in a way that typically requires authors to transfer their rights to their articles and books to the publishers who disseminate them. Publishers may seek to maximize their own returns by exploiting both the intrinsic value of the content they provide and their inherent monopoly over these products. The impact of copyright monopolies was mitigated, in the analog era, by purchasers' ability to loan physical objects. Restrictive digital licenses have led to very different results in the electronic marketplace, in which contracts trump copyright. The scholars who submit manuscripts for publication also work at a distance from the librarians who pay the bills. Prices and pricing policies are rarely visible across the academic community as a whole.

Some e-journal vendors have built up large portfolios that allow economies of scale. Additional price escalation often follows.[6] Big Deal arrangements push the model further as vendors bundle content to provide leveraged access to many journals at so-called special package prices. These packages carry drawbacks as well as benefits. The dynamics of entry and exit are problematic because a subscriber's initial Big Deal expenditure becomes its price floor for renewals. Contract terms often limit cancellations, title swaps, and other adjustments. Package prices climb at rates that can exceed both inflation and budget growth, distorting library budgets, collections, and programs. Some major vendors report profit margins as high as 30 to 40 percent, levels that most librarians find excessive.[7] Finally, librarians traditionally have built collections by applying policy prescriptions to guide their acquisitions. When both important and marginal materials are tied together, selection becomes moot. Even so, Big Deal costs typically fall below those of individual subscriptions (at dubiously determined list prices) to even a fraction of package offerings. The model is resistant to change.

Aggregation and bundling are obvious ways for publishers to extract value from the content they control. Aggregators also can explore diversified price models and marketing options to reach new customers. Offerings are thus targeted in terms of audience, functionality, usage rights, preservation options, and payment arrangements. Packages also can be presented as large databases of individual articles independent of the journals in which they appear, or with a focus on the journals themselves as both markers and makers of scholarly fields.

Any journal package faces practical thresholds of size and coherence, though additional enhancements can elevate these limits. Value-added features, which work across all the materials within a package, allow users to do things that would otherwise be tedious or difficult. The content, thus energized, becomes more useful and visible, permitting premium prices. Platforms, interfaces, and tools all contribute to functionality. Platforms combine disparate content into relatively seamless wholes. Interfaces, whether proprietary or freely available, allow users to discover and use materials in ways that are effective and appealing. Tools, whatever their origin, include features that can work across an entire interface and applications to support specific tasks. Well-designed delivery systems carry significant value of their own and strengthen the utility of prepackaged digital content.

Peer-reviewed e-journals, regardless of cost, conform well to the academy's exacting protocols for scholarly exchange. Community-inflected publishing ventures like Project Muse (http://muse.jhu.edu) or Project Euclid (http://projecteuclid.org) (for mathematics) can provide economies of scale without vendor profiteering. Nonetheless, journal concentration and exploitative pricing are utterly unsurprising consequences in high-return areas, such as science, technology, and medicine, and in professional fields, like law and business. Serials in the social sciences, some foreign-language offerings, and e-books are now being marketed with the same model—and with pricing expectations that similarly range from the ambitious to the downright predatory.[8] Libraries, too often caught between content providers and local users, are limited in what they can do.

Restoring the Balance

Research libraries, caught up in asymmetrical relationships with content aggregators, aspire to more balanced terms of power and trade. Both libraries and vendors are exploring new business and pricing models. At one end of the scale, prices might simply reflect production costs, perhaps also allowing a modest surplus. Other metrics of value look toward usage, costs and benefits, impact (by any number of measures), and return on investment. Some calculations anticipate comparisons across packages and between the titles within each one. Other approaches at once seek to educate users about journal costs and to reduce library expenses. Co-payments for article downloads, for example, might heighten user awareness and limit unnecessary usage, mimicking the charge-back logic that has helped to control the demand for "free" medical services.[9]

These and other proposals have had little effect to date. Different analysts, among them some who report to publishers and aggregators, can draw strikingly different conclusions from the same facts. For example, Ware and Mabe (writing for the International Association of Scientific, Technical and Medical Publishers) conclude that "the value for money that the Big Deal and similar licenses have brought, has largely contributed to the ending of the serials crisis, though that is not to say that the issue of journals cancellation has gone away."[10] Restraint-of-trade and anti-trust anxieties undercut calls for community mobilization. Concerted action between libraries and scholars has instead focused on consortial negotiations and open access (OA) publishing.

Library consortia, by combining their members' buying power and purchasing demand, aspire to secure favorable prices and license terms. Groups like the NorthEast Research Libraries consortium (www.library.yale.edu/NERLpublic) and the California Digital Library (www.cdlib.org) also have developed model licenses, standards for product performance, and best practices in areas like preservation. Vendors and libraries generally concur that consolidated operations for marketing, licensing, and invoicing reduce supplier overhead and therefore justify price breaks. Consortia may secure additional discounts, although data are sparse and cause-and-effect is difficult to assess. The internal workings of consortia reinforce the grounds for doubt. These bodies are instruments of their members' collective will, but also are beholden to each participant's priorities and claims. Group decisions are susceptible to lowest-common-denominator, weak-link-in-chain, and divide-and-conquer distortions. Consortia, in their current form, may be equivocal instruments of collective resolve.

OA publishing makes content freely available to all, without charge, through several increasingly complex forms. Many scholars post their research results to personal or group websites. Institutional and disciplinary repositories impose more robust controls through formal criteria for inclusion, standardized metadata, and purpose-specific software for deposit, retrieval, and retention. A growing number of universities, in some cases led by their faculty, require faculty members to deposit their articles in institutional repositories, usually with opt-out provisions. Funding agencies like Britain's Wellcome Trust (www.wellcome.ac.uk) similarly stipulate that publications resulting from the research they support be deposited in OA repositories. Researchers' participation in these repositories nonetheless varies widely, not least because of disciplinary cultures and deposit protocols. The high energy physics community, for instance, has been particularly successful with its arXiv preprint server (http://arxiv.org). Ongoing costs can be significant, as arXiv shows in its request for voluntary contributions.[11]

Personal websites and institutional and disciplinary repositories typically contain final prepublication drafts of articles and research reports; this is the realm of "green" OA. (Green OA is provided when an author publishes in a journal and then self-archives his or her contribution in an institutional repository or on some other OA website.)

Manuscripts will be modified during the editorial process, so version control is an ongoing concern. Most OA repositories also treat the items they hold as discrete and self-contained atoms of scholarship. While each such unit can be discovered and harvested by search engines, broader contextualization goes by the board. Tools to facilitate aggregation or provide added functionality have been slow to emerge within this diffuse environment.

Peer-reviewed OA journals, which compose the domain of "gold" OA in which publishers themselves make their products freely available, have provoked sharp debate as to sustainability and impact. Production costs, for example, cannot simply be wished away. Some OA journals rely on author fees that can exceed $1,500 per article, while others look to income from advertising, endowments, and third-party subsidies.[12] Groundbreaking projects like the high energy physics community's SCOAP[3] (Sponsoring Consortium for Open Access Publishing in Particle Physics) (http://scoap3.org) initiative are pursuing other approaches, in this case through funding pledges for an OA buy-out of core journals now supported by traditional subscriptions.

OA publishing provides the most promising (and, many believe, most affordable) alternative to the dominant subscription model. Some OA enthusiasts anticipate an eventual tipping point at which the volume, visibility, and reputation of these publications will prevail. This optimistic vision, however, may focus too narrowly on individual journals and the articles they contain. OA implementations lag far behind in bundling content and in developing the interface enhancements and value-added tools that commercial aggregators have successfully put into place. Moreover, OA may be too eagerly awaited as the key to a sustainable system of scholarly communication, when the real challenges lie in the larger structure of the academic enterprise.

Our model for scholarly communication took shape around analog publications. Today's capacities for combination, aggregation, and segmentation reflect the abundance of the digital universe. The intrusion of market economics in a knowledge-centered realm whose values and goals are largely distinct has resulted in content monopolies and price distortions. These economic pressures contribute to the sharply divergent approaches to collections, content, and services that now characterize our research libraries.

Paradigm Lost: Library Collections in a Fractured Landscape

While collections used to drive the research library agenda, this shared paradigm has lost its power. Local control of content is no longer the sole means to support research and learning, and building collections may not even be the best way to proceed. Three general approaches to collections and content are now taking shape, variously focusing on users and their needs, value propositions, and comprehensive holdings as ends in themselves. Both vendors and librarians are setting these terms. Collections cooperation also is on the rise, at least partly in response to cyclically tight budgets. The cloud's boundless capabilities and easy allure further affect our perceptions. Taken together, these elements have created an environment that may no longer cohere.

A growing and vocal contingent of research libraries is experimenting with patron-driven acquisitions.[13] Most mainstream trade and scholarly publications, by now produced and stored digitally, can be available on demand. Acquisitions that were once regarded as urgent and inevitable therefore can be deferred until a user requests them. A logical next step is to shift selection responsibilities from librarian go-betweens to users themselves. Patron-driven acquisitions align purchases with actual needs, allowing libraries to fulfill measurable service goals while minimizing putative losses for materials that are never consulted. Speed, responsiveness, and user agency bespeak nimble organizations that have adapted to our swiftly changing circumstances.

Other outcomes may prove less rosy. Patron-driven acquisitions can result in very similar and perhaps similarly shallow collections across different libraries. (Collections homogeneity also is a concern with other selection models.) User preferences can be manipulated. Reduced acquisitions, a celebrated corollary to tightly targeted purchases, may further undercut sales for academic and specialized publishers, with damaging spillover effects for individual researchers and also the broader system of scholarly communication. Finally, patron-driven acquisitions presume large, predefined pools of candidate publications from which choices can be made. Research libraries typically have ventured beyond readily identified materials to pursue niche publications and primary sources, categories omitted from most demand-driven scenarios.

Another collections strategy focuses on returns on investments. Several studies have correlated libraries' Big Deal subscriptions with the high level of package-based citations that then characterize their institutions' successful grant proposals.[14] These exercises are compromised by vendor sponsorship and by their inference of causality from correlation. Nonetheless, the broader proposition that well-spent collections budgets should generate measurable returns seems entirely plausible. Finding models and metrics to pursue high-value content also makes sense.

Finally, broad coverage and unique acquisitions remain dear to some large libraries. These repositories, like all others, recognize that they will never capture everything that their users might desire. They also struggle with digital primary sources and raw data, categories whose nature and scale currently exceed both imaginations and capacities. Inclusive coverage nonetheless remains the goal.

Some libraries partake of all three models as they focus in turn on undergraduate learning, professional programs, and original research. Despite their differences, these approaches embody some common principles. Academic libraries today place users at center stage, albeit in a tableau complicated by the varied requirements of faculty members, graduate or professional students, and undergraduates; and the interplays between current and future needs, and between local and off-site constituencies. Research libraries also participate in cooperative initiatives that, together, may help compensate for reduced local acquisitions.

Research libraries have a long history of cooperation around collections and content. Most participated in such emblematic projects of the analog era as the National Union Catalog to facilitate interlibrary loan; the Farmington Plan to ensure comprehensive coverage of foreign imprints; and the Association of Research Libraries (ARL) to establish and then energize their community. Geographic proximity allowed a second set of collaborations, for example between Duke and the University of North Carolina and between Berkeley and Stanford. A third approach constructed deep, specialized collections in particular fields, for instance through area studies microfilm projects.

Cooperation remains a community preoccupation, but its contours have changed. Venerable umbrella organizations like the ARL and the Center for Research Libraries (www.crl.edu), plus newcomers like the Digital Library Federation (www.diglib.org) and the HathiTrust (www.hathitrust.org), continue to feature both strategic leadership and practical initiatives. Modest resources, however, limit their potential as community counterweights to marketplace behemoths. Research libraries are likewise cultivating intense interdependencies; a good recent example is the "radical collaboration" between Columbia and Cornell.[15] Many cooperative programs still focus on enhanced access to elusive primary sources and other unusual materials.

Today's most innovative cooperation, however, may be occurring at the regional level. For example, and as a complement to its initial focus on unmediated and expedited book deliveries between member libraries, the Borrow Direct group is exploring far-reaching resource sharing and collection complementarities between its northeastern members.[16] The Association of Southeastern Research Libraries (www.aserl.org) has launched a project to rationalize holdings of U.S. Federal Depository materials, and the Committee on Institutional Cooperation (CIC) Center for Library Initiatives (www.cic.net/Home/Projects/Library) and the Western Regional Storage Trust (WEST) (www.cdlib.org/west) are grappling with the cooperative preservation of digital and physical holdings. These activities, at once ambitious and limited, bespeak a promising though somewhat disjointed response to the community's multilayered collections agenda. Also, the total resources now available for these cooperative programs are only a fraction of the amount that would reset research libraries' collective capacity.

In a related area, commercial products are filling gaps that research libraries might have addressed on their own. Past partnerships between libraries and microfilming agencies prefigured today's offerings of prepackaged, thematically coherent digital sets of primary sources. Libraries, generally drawing only from their own collections, have created some similar but freely available aggregations. All of these packages allow students to confront original sources while relieving libraries of the need to pursue extensive, specialized acquisitions on their own. Such uncomplicated access to sometimes-obscure content, however, may carry unintended consequences for both libraries and learners, not least as overall coverage diminishes and as inevitably partial holdings are presumed to provide balanced representations of some larger whole.

Research library needs also are being addressed through products, like JSTOR (www.jstor.org), created by nonprofit organizations that seek sustainable benefits for the entire academic community. Some such initiatives also address local needs that are beyond local capacities. The emergent force in today's information domain, however, is the cloud and cloud computing. Google's direct control of content, search engine predominance, and grasp and ambition make it the lead player. Libraries have yet to articulate the principles or mobilize the means to have a comparable impact.

Research libraries are redefining their collections-related strategies and goals. Higher education is itself increasingly diverse. Today's many library agendas allow experiments and innovations that were hard to launch when the collections paradigm reigned supreme. The community's turn from monolithic aspirations for comprehensive collections and content nonetheless raises questions at the same time that it creates opportunities.

Emergent Complexity, Creative Destruction, and Research Libraries in the Digital Age

Ongoing changes in information, scholarship, and institutional structures, leavened by market dynamics, pose similar challenges to all research libraries. Their responses then reflect local priorities, perspectives, and resources. Some forms of cooperation have become more prominent. Other measures, including the current fixation on patron-driven acquisitions, are still taking shape. "Emergent complexity," a general feature of evolving systems, typically includes outcomes that are unexpected as well as intentional. Careful analysis and anticipatory action can improve the chances for desirable results. The following considerations will affect all research libraries' strategies and plans, whatever they do. Four elements relate to the economics of collections

and content; the other five involve more general aspects of theory and service. All should inform discussion.

Aggregation

Commercial e-journal vendors have led the way in combining digital content into the large bundles that permit economies of scale and produce scale effects for tools and applications. The components that make up a package, typically articles and journals, become more visible as the package itself becomes more robust. Research libraries need to develop their own aggregated content and also to go beyond single-institution silos. Digitized holdings, websites, departmental publications, research findings, institutional repositories, and data sets could all form part of these new aggregations.

Value-Added Features

Commercial aggregators again have taken the lead in providing features and functionality to enhance their products. Proprietary platforms, however, can reinforce dependent relationships and dispersed content. The research library and academic communities, sometimes with foundation support, are already pursuing standards-based, open-source systems and tools.[17] These efforts require additional infusions of cash, staff, energy, and resolve.

Versioning and Segmentation

Information vendors often customize their products for specific consumer niches through different combinations of features and content. But research libraries' offerings, for example their topical collections of digitized local holdings, typically presume that one size fits all. Different versions adapted to multiple audiences and needs—secondary schools, undergraduates, researchers—could attract new constituencies. Wider recognition and new revenue streams might follow.

Marketing

Many research libraries are uncertain when it comes to publicizing and projecting their collections and capabilities. Simply branding what the library provides is one basic step. In a larger sense, marketing is a strategic process to align an agency's services and goals with a community's needs, thereby enabling an ever more vital relationship. Research libraries' confused sense of purpose, within and beyond the collections realm, contributes to mixed messages, fragmented actions, and spotty support. Consensus and clarity will serve us well.

Time-Value and Velocity

The relationships between user behaviors, the availability of information, and time are complicated and poorly understood. Digital technologies have redefined timeliness as it applies to teaching, research, and scholarly communication. With instant gratification the norm, even brief delays create frustration. The velocity of information is more complex. As with money, the impact of information increases with its use; rapid turnover and dense transactions create and embody value. Velocity also is associated with turbulence in ways we do not fully grasp. The velocity of information needs further analysis in terms of its interplay with selection strategies, collections goals, and value-added features.

Agency

Research libraries played a determinative role in building the hard-copy collections that historically both allowed and limited local scholarship. Today's abundance of freely available digital content has democratized access and broadened agency. The library's continuing centrality in creating, shaping, and interpreting this environment is no longer clear. We need to highlight and hone the functions that libraries continue to fulfill.

Cutbacks and Communications

Local priorities and constraints are pushing some research libraries toward measures that may diminish the larger community. Unique materials always have been lost because of negligence, and in some cases, design. Ongoing reductions make prospective collections gaps ever more likely, with ripple effects that may undermine related services ranging from cataloging to interlibrary loan. The free-rider imbalances that may arise as patron-driven collections increase the community's dependence on a few continuing strong collections also may require new cost modeling beyond the transaction-driven charge-backs now in place. All of these perhaps unintended consequences could be mitigated through mechanisms for libraries to communicate their cutbacks and policy shifts and then to allow others to respond.

Description and Organization

Libraries provide organized access to recorded knowledge through standardized descriptions, controlled terminologies, indexes to holdings, and bibliographies and guides. Even as search engines and algorithms continue to improve, many observers insist that we will always need structured metadata and normalized descriptions. New tools may automatically extract and compile the necessary data elements, though some argue that only human interventions can ensure success. The issues are both philosophical and practical, and may play out differently across different fields. The approaches that prevail will fundamentally affect the nature and the scale of library-based collections and content.

Preservation

Most research libraries are focusing on user services. Preservation, which also is associated with the waning realm of collections, is a less immediate concern. Even so, conversations are already underway to harness emerging technologies and cooperative capabilities for community-level action.[18] Some collections-heavy repositories also may end up carrying a disproportionate share of the burden. Service-level agreements for preservation, models to quantify all the associated costs, and adequate revenue streams require both careful analysis and full participation.

Preservation raises other questions as well. Researchers, archivists, and librarians warn of an incomplete record of scholarship and lost primary evidence as digital files vanish forever. We still do not fully understand the basic infrastructure that is needed to preserve all varieties of digital information. Models to identify the material that we should preserve are likewise not yet in place, and the resources that would allow us to move ahead are huge. Most libraries are acquiring ever smaller segments of a burgeoning output of both analog and digital information. With so much content beyond our grasp, significant gaps are inevitable.

Implication of the Considerations

These nine elements of these collections and content landscape, and perhaps others as well, require focused attention as we reassess our circumstances and goals. The most obvious features of today's transformed information landscape, among them vast content available without mediation, costly bundles and packages, and new business and operational models for scholarly communication, easily draw attention toward symptoms rather than causes, and coping mechanisms rather than cures. Many libraries also are taking action only on the basis of their own perceptions and needs. A broader perspective is in order.

Economists identify "creative destruction" as an inherent, impersonal, and typically disruptive feature of change. Scientists recognize extinction, along with emergent complexity, as a byproduct of systems in flux. Shared analysis and coordinated plans can help ensure that the essential functions of research libraries—identifying, providing access to, organizing, delivering, and preserving recorded knowledge—continue to be fulfilled.

Conclusion

Collections no longer lie at the center of research library operations and goals, even as academic communities focus ever more inclusively on knowledge and information. This essay has explored some of the salient transformations and pressures that underlie these paradoxical shifts. Our increasingly digital environment, insistence on educational outcomes and accountability, and appropriately supportive library services are all in play. The sometimes simplistic lure of cooperation likewise affects the landscape. Each research library then faces its own economic and organizational environment. In the absence of common strategic perspectives and priorities, libraries engage in increasingly diverse, sometimes idiosyncratic, local responses. The consequences are highly uncertain for a community that has historically embraced joint actions to bolster an academic enterprise that requires open communication and readily accessible knowledge. Several specific considerations, some of them primarily economic in nature—although others are less tangible—will help inform research libraries, separately and as a community, as they continue to support research and learning.

References and Notes

1. Research libraries make up an elusive category. The principles and criteria for membership in the Association of Research Libraries, which includes about 120 repositories in the United States and Canada, provide a working definition. See Association of Research Libraries, About ARL, Membership, Principles of Membership in the Association of Research Libraries, http://www.arl.org/arl/membership/qualprin.shtml (accessed Feb. 15, 2011).
2. Thomas S. Kuhn, *The Structure of Scientific Revolutions*, 3rd ed. (Chicago: Univ. of Chicago Pr., 1996).
3. Project Information Literacy: A Large-Scale Study about Early Adults and their Research Habits, www.projectinfolit.org (accessed Mar. 21, 2011) documents a massive and continuing effort at the University of Washington's Information School. See also John Palfrey and Urs Gasser, *Born Digital: Understanding the First Generation of Digital Natives* (New York: Basic Books, 2008).
4. Chris Anderson, "The Long Tail," *Wired* 12, no. 10 (Oct. 2004), www.wired.com/wired/archive/12.10/tail.html (accessed Mar. 21, 2011) provides a classic modern analysis of uneven distributions of demand and use across large sets, specifically suggesting that digital technologies can facilitate marketing and access that would otherwise be prohibitively expensive.
5. The literature on serials pricing issues is immense. An excellent, exhaustive starting point is Charles W. Bailey Jr., *Scholarly Electronic Publishing Bibliography*, vers. 79 (Dec. 13, 2010), www.digital-scholarship.org/sepb/sepb.html (accessed Feb. 15, 2011).
6. The extraordinary price increases announced by the Nature Publishing Group precipitated highly public exchanges with the University of California during the fall of 2010. See, for example, Jennifer Howard, "Nature Publishing Group Defends Its Price Increase for U. of California," *Chronicle of Higher Education* (June 9, 2010), http://chronicle.com/article/Nature-Publishing-Group/65848 (accessed Feb. 15, 2011); and Laine Farley, Richard A. Schneider, and Brian E.C. Schottlaender, "Response from the University of California to the Public Statement from Nature Publishing Group Regarding Subscription Renewals at the California

Digital Library," June 10, 2010, http://osc.universityofcalifornia.edu/news/UC_Response_to_Nature_Publishing_Group.pdf (accessed Feb. 15, 2011).

7. Glenn S. McGuigan and Robert D. Russell, "The Business of Academic Publishing: A Strategic Analysis of the Academic Journal Publishing Industry and its Impact on the Future of Scholarly Publishing," *Electronic Journal of Academic & Special Librarianship* 9, no. 3 (Winter 2008), http://southernlibrarianship.icaap.org/content/v09n03/mcguigan_g01.html (accessed Feb. 15, 2011).
8. The Italian aggregator Fabrizio Serra has become notorious for acquiring scholarly titles in the humanities and social sciences whose prices then explode. See Claudio Giunta, "Quanto (ci) costa la cultura accademica," *La rivista dei libri* (Feb. 2010), www.claudiogiunta.it/wp-content/uploads/2010/02/quanto-ci-costa-leditoria-accademica_.pdf (accessed Mar. 21, 2011).
9. California Institute of Technology's co-pay approach to providing articles in journals to which CalTech does not subscribe is described in Caltech Library Services: How to Find a Journal Article, http://libguides.caltech.edu/howtofindajournalarticle (accessed Feb. 15, 2011). The library also provides subscription access to many journals. See also Stuart M. Shieber, "Equity for Open-Access Journal Publishing," *PLoS Biology* 7, no. 8 (Aug. 2009): 1–3, http://www.plosbiology.org/article/info:doi%2F10.1371%2Fjournal.pbio.1000165 (accessed Feb. 15, 2011).
10. Mark Ware and Michael Mabe, *The STM Report: An Overview of Scientific and Scholarly Journal Publishing* (Oxford, UK: International Association of Scientific, Technical and Medical Publishers, Sept. 2009): 62, www.stm-assoc.org/news.php?id=255 (accessed Feb. 15, 2011).
11. Cornell University Library, arXiv Support, http://arxiv.org/help/support/ (accessed Feb. 15, 2011).
12. Many OA journals (some calculations put the number as high as 75 percent) charge no fees. Additional data compiled by Stuart Shieber and others suggest that OA charges, as applied by journals with Institute for Scientific Information (ISI) impact factors, average around $1,100. PLoS One carries a fee of $1,350. See Stuart Shieber, "What Percentage of Open-Access Journals Charge Publications Fees?" online posting, May 29, 2009, The Occasional Pamphlet, http://blogs.law.harvard.edu/pamphlet/2009/05/29/what-percentage-of-open-access-journals-charge-publication-fees (accessed Feb. 15, 2011); and his "Is Open-Access Journal Publishing a Vanity Publishing Industry?" online posting, The Occasional Pamphlet, Oct. 16, 2009, http://blogs.law.harvard.edu/pamphlet/2009/10/16/is-open-access-publishing-a-vanity-publishing-industry (accessed Feb. 15, 2011). See also Raym Crow, *Income Models for Open Access: An Overview of Current Practice* (Washington, D.C.: Scholarly Publishing & Academic Resources Coalition, Sept., 2009), www.arl.org/sparc/bm~doc/incomemodels_v1.pdf (accessed Mar. 21, 2011).
13. The combined issue of *Collection Management* 35, no. 3/4 (2010) is devoted to patron-driven acquisitions. Many conference sessions also have considered this topic, for example the 2010 Charleston Conference program, www.katina.info/conference/program.php (accessed Feb. 15, 2011) and the American Library Association's 2011 Midwinter Meeting.
14. Judy Luther, *University Investment in the Library: What's the Return? A Case Study at the University of Illinois at Urbana-Champaign*, White Paper no. 1 (San Diego, Calif.: Library Connect Editorial Office, Elsevier, 2008), http://libraryconnect.elsevier.com/whitepapers/lcwp0101.pdf (accessed Feb. 15, 2011); and the follow-up by Carol Tenopir et al., *University Investment in the Library, Part II: An International Study of the Library's Value to the Grants Process*, White Paper (unnumbered) (San Diego, Calif.: Elsevier, 2010), http://libraryconnect.elsevier.com/whitepapers/roi2/2010-06-whitepaper-roi2.pdf (accessed Mar. 21, 2011). Other value metrics include those being developed at the California Digital Library and mentioned in Ivy Anderson, "Challenges to Licensing from Some Publishers," Jan. 27, 2011, www.cdlib.org/services/collections/current/challenges.html (accessed March 21, 2011).
15. "Radical collaboration" is proclaimed in the joint statement of University Librarians Anne Kenney (Cornell University) and Jim Neal (Columbia University) on the 2CUL homepage, http://2cul.org, and explicated with additional detail throughout the site (accessed Feb. 15, 2011).
16. Borrow Direct is a rapid book request and delivery system that enables faculty, staff, and students at Brown, Columbia, Cornell, Dartmouth, University of Pennsylvania, Princeton, Yale, Harvard, and MIT to search the combined library catalogs of these libraries and directly request expedited delivery of circulating items.
17. See, for example, descriptions of foundation-supported, open-source systems and tools such as Project Bamboo (http://projectbamboo.org) and Kuali Open Library Environment (OLE), (http://www.kuali.org/ole) (accessed Feb. 15, 2011).
18. See, for example, Robert H. Kieft and Lizanne Payne, "A Nation-Wide Planning Framework for Large-Scale Collaboration on Legacy Print Monograph Collections," *Collaborative Librarianship* 2, no. 4 (2010): 229–33, www.collaborativelibrarianship.org/index.php/jocl/article/view/119/77 (accessed Feb. 15, 2011); and Constance Malpas, *Cloud-Sourcing Research Collections: Managing Print in the Mass-Digitized Library Environment* (Dublin, Ohio: OCLC Research, 2011), www.oclc.org/research/publications/library/2011/2011-01.pdf (accessed Feb. 15, 2011).