

# Rethinking Research Library Collections

## A Policy Framework for Straitened Times, and Beyond

By Dan Hazen

*Academic and research libraries today confront daunting financial pressures. Their faltering budgets also compound an intensifying existential crisis resulting from profound shifts in information, scholarship, technology, and academic organizations. The purposes of collections are particularly uncertain in this radically fluid context. Analyzing the most salient elements in today's collections landscape can help to frame the guiding principles that will inform adaptive new approaches to collections and content.*

Research libraries today contend with shrinking budgets that compound a long-standing structural mismatch between available resources and community expectations. The broader landscapes of information, scholarship, technology, and academic organizations also are in flux. The community's collections strategies must therefore adapt to a radically fluid context that is brimming with both opportunities and demands. This essay describes some key elements in today's collections landscape and also offers a simple model for information types and their uses. The framework in turn suggests a set of principles to inform a redirected strategy for collections and content.

### Universities, Information, and Library Collections: An Environmental Scan

#### The Information Landscape: Continuity and Change

##### Information

The supply of information resources has mushroomed across all formats. Emerging countries and also traditional publishing centers are producing more than ever before. Recession and deflation may mitigate these trends, and some categories of publications—print newspapers are a likely example—may decline or even disappear. Nonetheless, predictions that hard copy publications will soon be overwhelmed by an avalanche of electronic resources are emphatically premature, particularly in the developing world and other areas of emergent modernity. Analog materials remain both prevalent and indispensable as the digital explosion continues apace.

Despite the persistence of print, large-scale digitization is transforming the library world. The scale of the resources available via Internet search engines

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

Submitted November 9, 2009; accepted November 19, 2009, pending modest revision. Revised and submitted December 4, 2009, and accepted for publication.

This essay is based on a longer discussion paper the author prepared for the Harvard College Library in the spring of 2009.

substantially exceeds that of any single research library, and also of the research library community as a whole. The immense range of virtual materials now at hand has in turn undermined the quantitative measures by which we have traditionally judged our holdings. Libraries and librarians, along with other agencies devoted to our intellectual and cultural heritage, are experiencing a dual crisis of purpose and identity. Most students and scholars perceive less cause for concern.

The day's expanding array of information resources, in all formats, is complemented by intense price pressures that consistently outpace inflation. Publisher conglomerates, which already wield oligopolistic control over the scientific, technical, and medical information universe, are now expanding into other market segments. Greater outputs of increasingly expensive published materials characterize developing regions throughout the world. The weakened dollar, which lost about one-third of its value relative to both the Euro and the Pound between 2000 and 2008, compounds the challenge for libraries with heavily international collections. Information may want to be free, but it also is a commodity—and scholarly resources are in thrall to the marketplace.

#### Scholarship

Research and teaching continue to evolve. Until recently, pedagogical models and research strategies privileged the “core” or “canonical” writers and sources. In most fields, scholarship was considered an orderly and necessarily cumulative enterprise in which new inquiries both relied and built upon the earlier studies that composed the scholarly record. The bulk of research and learning was confined within rigid disciplinary boundaries, with each field claiming its own foundational literature and a unique suite of closely aligned methodologies.

Today's expectations are profoundly different. Cross-disciplinary inquiry, participatory learning, an obsession with primary resources and original documentation in all formats, and hybrid methodologies are increasingly the norm. The record of scholarship, while still important, has in many fields become less central. Multimedia research products, scholarship that relies upon massive and remotely hosted data sets, and team-based inquiry are other features of this emerging panorama. The appropriate locus of support for these resources is not yet clear.

#### Technology

In the past, libraries tended to acquire and warehouse hard-copy materials as passive objects for students and scholars to ferret out and then interpret on their own. Digital resources, by contrast, are energized from the start. The catchphrase

“If it isn't on Google, it doesn't exist” only begins to capture this dynamism. For a simple example, keyword searches across JSTOR can lead researchers to sources that would have remained invisible in a context limited by the traditional apparatus of field-specific bibliographies, indexes, and abstracts. On a more mechanical but likewise transformative level, linked footnotes allow the seamless pursuit of citation threads that would be unrealistic to track through physically dispersed books and journals. Mash-ups and other digitally recombinant possibilities encourage projects that transcend an exclusively textual framework. Libraries are now called to help users by contextualizing all of these energized resources within a broadly activated system of information, tools, and expert staff. Integrating our deep stores of analog holdings into this high-energy electronic network remains a central challenge.

Digital information predominates within some fields and is increasingly prevalent across the board. Digital technologies also are affecting scholarly inquiry and output as well as teaching and learning. Large data sets—numerical data, text corpora, image banks, etc.—invite structured inquiries across masses of information on a scale that can easily exceed the capabilities of any single institution. Tools for analysis, manipulation, and visualization may best be developed as community efforts. The “cloud” is becoming the locus for more and more data and applications in contrast to past models, which had readily identified developers and sites. “Community engagement” is likewise the byword for social networking initiatives, whose shorthand is the panoply of Web 2.0 products and services. Many new approaches to teaching and learning similarly rely upon open-source collaborative and participatory tools.

We do not yet understand the scholarly significance of large swaths of the digital universe. Blogs are often compared to diaries; e-mails are likened to letters and memos. The analogies are not only imperfect, but they may also complicate our decisions about what we need to capture and preserve. More familiar products like learning objects and computer software can be difficult to assess. Websites are typically dynamic and multilayered, requiring thoughtful protocols to determine what to retain. Social networking spaces are again unfamiliar. Instant Messages and cell phone videos pose challenges of their own. Scholars, users, creators, technologists, librarians, and digital objects themselves all have roles to play in clarifying our possibilities and needs.

One critical, perturbing, and unresolved element within the electronic universe concerns the requisites for preservation, which are most effectively addressed at the moment of digital conception. Today's technologies for access control—digital rights management, streaming systems, legal and contractual limitations, and so on—often work at cross purposes to permanence. Consensus-based regimes to ensure digital persistence are far from certain.

### Organizations and Institutions

The academic world is moving beyond structures defined primarily by discipline. Newly minted centers, institutes, programs, and initiatives today provide homes for interdisciplinary scholarship even as traditional departments remain strong.

The scholarly community also is affected by other kinds of structures and constraints. Intellectual property regimes channel access to and uses of many information resources. Google is an archetype for the commercial players that now occupy an expanding and disquieting space within the realms of information and academia. Traditional higher education is itself struggling against intense financial pressures, with for-profit institutions promoting an essentially distinct vocational model.

Cooperative arrangements and consortia are further reshaping the institutional environment. Economies of scale, aggregated expertise, new synergies and unexpected opportunities, and strengthened political coalitions and operational capacities are among the potential benefits. Local autonomy is less possible or desirable than ever—even as institutional competition remains a hallmark of American higher education.

### Modeling Information, Collections, and Content

Academic institutions create and also consume information. Libraries play a critical role within this ecology as they ensure the community's continuing access to the information resources that sustain research and learning. Conceptual frameworks, as well as practical tools, enable libraries to understand and then manage the torrents of information that now overflow the landscape. The following heuristic model has been created to help clarify our options. This model asserts that information resources in all forms and formats, whether viewed individually or in broader groupings, can be clumped into four ideal categories that reflect their academic uses as well as their origins: core resources and curricular support, the record of scholarship, primary resources, and data. Libraries, along with scholarly disciplines, departments and programs, and individual students and scholars, play critical roles in enacting this classification.

#### Core Resources and Curricular Support

All academic libraries provide the basic bibliographies and reference works, reading list materials, foundational literatures, and other core sources that are required for teaching and learning. Curricular support is a fundamental activity for every college and university library. Local definitions of each field's core resources also tend to carry across from one institution to the next.

### The Record of Scholarship

Academic libraries in institutions that support original research and advanced study further aspire to capture some or all of the record of scholarship. New studies in many fields are framed within a broader context of ongoing inquiry, as manifest in the scholarly record. Holdings that recapitulate this record thus remain critical in sustaining the cumulative process of creating new knowledge. This category includes the published outputs of colleges and universities, think tanks and scholarly societies, commercial laboratories and trade organizations, academies and associations, specialized agencies, and ad hoc research groups. A particular library's collecting appetite may vary within this large realm—perhaps only American university press publications; a multinational, multilingual sampler; or (at least in theory) exhaustive coverage. Levels of coverage can also vary among fields. Electronic publishing and new access technologies may mitigate the need for comprehensive local collections of the scholarly record by providing alternate ways to locate and use these resources.

The record of scholarship manifests itself above all in books and journals. Scholarly journals, which are important in all fields, make up the primary vehicle for validating new findings in science and technology. Market dynamics are pushing serials toward digital formats, through which they can also be disseminated at multiple levels of aggregation (bundled journal packages, individual serial titles, and specific articles)—always within a context of escalating costs. Scholarly monographs are then particularly central to the humanities. Despite experiments in electronic publishing and forecasts of ubiquitous print on demand, these materials are currently at risk.

#### Primary Resources

This immense third category comprises all organized human expression, or the full range of primary sources. These raw materials for scholarly work have become ever more eclectic. Many libraries have always pursued a broad range of noncanonical creative writing—novels, drama, poetry, and so on. Local and international newspapers, as well as government documents, are enduring mainstays as well. The scholarly record and synthetic works themselves serve as primary sources for researchers studying intellectual history and broader shifts in ways of thought. Rare book holdings and many special collections fall within this category as well.

Other primary sources have only more recently gained a place within the library (and scholarly) pantheon. Ephemera and grey literature, pamphlets, popular magazines, comic books, visual imagery, films and video, manuscript and archival collections, and sound recordings are all by now

accepted as legitimate collections categories. Websites, blogs, and other digital outlets, sometimes created as social endeavors and sometimes to represent a single perspective, are more recent additions. On a global scale, the gradually diminishing digital divide still affects information production and collections strategies across areas that differ in terms of affluence or openness.

#### Data

Unorganized or minimally structured raw data represent a category of information that we are only beginning to understand. Scholars' unprocessed laboratory notes and research transcripts—unruly file cabinets, boxes of scribbles and scrawls—provide a quaintly venerable and readily managed example. The realm of raw data has assumed greater importance as the research tools associated with “big science” drive more and more scholarship. Digital satellite imagery, DNA and genome sequences, remote sensing data, raw survey responses, meteorological measurements, and text and image corpora are among the data sets and data streams that now pose daunting challenges of capture, interpretation, and curation.

Each field's scholarship and teaching draw upon different blends of information from these four categories. Research in medieval studies, for example, relies upon an array of original sources and texts that is by now pretty much fixed, at least when compared to the endless tidal wave of new materials that inform scholarship in fields like film studies, chemistry, and political science. For medievalists, exhaustive access to contemporary scholarship is therefore essential. Even this ground, of course, is not entirely solid—for instance, the field's research has broadened beyond a fairly confined textual canon to include the evidence of archeology and material culture. The high energy physics community, by contrast, relies heavily on the almost comprehensive availability of research findings in the arXiv (the archive for electronic preprints of scientific papers hosted by Cornell University) server. Peer-reviewed journals then invest specific reports with validation and prestige. Vast streams of raw data, for instance those generated by CERN's Large Hadron Collider, are crucial as well. Benchmark monographs serve to recapitulate the field's state of the art at particular points in time.

The academy's lore depicts the library as the humanist's laboratory, implying that historical materials and primary sources are less central in other scholarly realms. We need a more nuanced understanding. For example, historic field surveys are indispensable for botanical and zoological research. Star maps and celestial observations from both past and present are essential for astronomers. Scholars' uses of noncurrent literature in disciplines like chemistry or physics may follow in the path of medical researchers, who

have fruitfully engaged in text mining across large sets of historic data and reports.

The academic uses of information resources are shifting, sometimes in unexpected ways, across all four categories. While discipline-specific research paradigms often remain important, more agile models for scholarship and inquiry also suggest more fluid approaches. Today's pedagogical models routinely require students to grapple with primary sources and special collections, as well as secondary works and synthetic texts. Just as individual resources have become energized in the current environment, so has the entire structure of information. Our traditional collecting expectations, which were far more static and staid, no longer serve us well.

#### The Changing Contexts and Expanding Scale of Collection Development

Models for information resources provide one potentially useful window into library collections and collecting. Another perspective focuses on the changing context within which our collections are now being built.

##### Collections of Record, Collections for Use

With a few exceptions, such as consciously duplicated core materials, reserve readings, and high-use recreational works, research libraries have sought to build collections that will persist through time. Carefully selected individual items, in their aggregate, make up definitive representations of the associated topics and fields. Libraries then care for these assemblages so that they will be permanently available. Creating and stewarding this patrimony constitutes a vocation of broad cultural consequence.

Looking to the future, research libraries will in some areas continue to build enduring collections of record. In others, they will settle for use-driven holdings while seeking neither comprehensive coverage nor long-term retention. The availability of digital surrogates or of remotely maintained archival copies may also affect local choices. Ideally, libraries will seek to ensure that some institution is providing ongoing preservation and care for everything they hold—but there may be instances in which current-use materials are acquired and discarded regardless of provisions for persistence. The continuum of curation will become more diverse.

##### From Collections to Collections and Content

Most academic libraries will continue to acquire both analog and digital materials for their on-site collections. However, their focus will expand ever more emphatically beyond acquisitions as they also provide access to intellectual content that

is leased rather than acquired, or to which they only point. Some libraries will likewise continue to create new, primarily digital, resources on their own. The increasing ubiquity and utility of highly diverse digital resources will require adjustments in all library operations. “Content”—a category that encompasses everything to which a library enjoys ready physical or digital access regardless of ownership status—is central to all that we do.

The diffuse knowledge that is embedded within and suffused throughout every university is a form of local content that most institutions have barely begun to tap. Energizing and leveraging this largely latent capacity is critical to the academy’s future. The process will most fruitfully engage faculty, staff members, and also the students whose research pilgrimages—mental and physical—foreshadow tomorrow’s scholarly agendas. Knowledge management will be a necessary element in our emerging content strategy.

#### Enlarging the Field: Partners and Players

All academic libraries are under intense financial pressure. The possibilities and the shared challenges associated with digital resources, the scale of today’s information needs, and examples of consortial achievements together make cooperation more appealing than ever before. The production of information resources, as well as conjoined consumption and processing, can become shared functions within a virtual environment.

Collective action may allow libraries to more fully shape both the landscape and the marketplace for electronic resources. Collections cooperation has traditionally emphasized the obscure, low-demand, sometimes expensive resources that can be shared between partners with minimal inconvenience to occasional local users. The compelling argument holds that shared physical resources made available through interlibrary loan can effectively reduce the need for redundant acquisitions at many different sites. Collecting scale, geographic and programmatic proximities, and resonances with other cultural institutions further shape the potential results. Structured acquisitions programs and streamlined processes for resource sharing have allowed limited progress in some relatively specific collections niches.

This model might now also be turned on its head as members of consortia together identify and arrange for digital access to core materials. Particularly in the electronic age, cooperative activities can cut across all four categories of collection resources. For large data sets, collaboration will be essential in building both infrastructure and tools because of the sheer scale of the task.

Collaborative action might encompass other dimensions as well. Libraries, archives, and museums are often co-located. They also share similar aspirations and missions. New opportunities for service and deeper complementarity

may be at hand. Research library cooperation has been most successful in focused efforts between groups of limited size, for example, intensive partnerships between two or three peer institutions, and relatively compact consortia such as the Committee on Institutional Cooperation or the California Digital Library. Cooperative initiatives that achieve enduring operational success seem to be bound by intractable limitations of organizational structure and scale, even in today’s technological age.

#### Libraries as Storehouses, Libraries as Tool Sheds

The mass of information resources now available on the Web, many of them free, is fundamentally changing the library community’s thinking about collections. High-quality and openly accessible scholarly resources—digitized maps and medieval manuscripts, books and journals, images from archives and art museums, music scores and sound recordings, and so on—can be found in staggering profusion without even considering the medium’s less scholarly emanations. Links to freely available digital content, metasearch capabilities that cut across products and platforms, and local aggregations of electronic resources, will all play a growing role in libraries’ collections and content strategies. This in turn will also reduce the physicality of library holdings and alter the functionalities of their spaces. But we need to go further.

Three aspects of Web-based content require close attention. First, the search engines that today allow users to find materials on the Web are neither transparent nor fully revealing of useful content in predictable ways. Google Scholar, for example, relies upon opaque search algorithms and relevance rankings that appear not to fully exploit the wealth of standards-based metadata that libraries routinely provide. But most libraries do little better, investing their cataloged resources with robust metadata that our discovery tools rarely handle well. Second, sources on the Web—whether websites themselves or the data, images, objects, and documents embedded within them—are notoriously unstable. Content is added, changed, and removed; links shift around and disappear. Scholarship relies on enduring access to constant content, a goal that remains elusive in the digital domain. Capture, curation, and digital preservation are all implicated in this conundrum. Third, dispersed and disparate Web content requires tools that can work across amalgamated sets of sources in predictable and repeatable ways. Some of the uses are well understood while others reflect a new realm of inquiry that includes text mining, pattern recognition, visualization, and simulation. The needs are perhaps most pressing around massive accumulations of raw data.

Libraries, working together and also with academics and information technologists, have an evolving role in creating

and supporting the tools that will enable students and scholars to take full advantage of the digital world. It is not yet clear whether lead roles can or should be preordained; arrangements that embody flexibility and contingency seem most likely to succeed.

#### Scarcity: Measure of Prestige or Consequence of Manipulation

Research libraries have traditionally built their reputations on the basis of their collection size and also the depth and breadth of their rare book holdings and their special collections. Scarce or unique artifacts, as well as uniquely comprehensive collections, remain primary measures of quality. Prestige based on both size and scarcity may diminish as large-scale digitization weakens the once obvious benefits of local ownership. The structural scarcity associated with rare artifacts is less compelling in a rich digital environment.

Paradoxically, our most coveted resources now include those digital materials whose uses are limited by contractual restrictions. Electronic gatekeepers can create scarcity (and also compromise long-term persistence) by manipulating license agreements and relying upon restrictive delivery technologies, even as the underlying resources could in theory be available without limit. “Scarcity,” in a traditional sense, reflects materials that are physically rare or unique. Today’s environment adds in the artificial scarcity created through restrictive manipulations of the digital marketplace.

#### Authorship and Authority

Academic libraries have historically served as custodians for carefully selected, authoritative information. Library holdings were then taken to embody the highest standards of analytical and methodological rigor. The weighty bound tomes associated with research libraries and traditional scholarship carried their own aura of permanence and security. Norms for careful reading and measured scholarly discourse further suggested prudence, stability, confidence, and authority. Deeply embedded synergies between artifact and text played an essential role in research and teaching.

Our excursions into the exuberantly expressive realm of primary resources have effectively destroyed these presumptions. All manner of deliberately ephemeral products circulate at high velocity, undermining anyone’s attempts to delimit agency, define a “canon,” or codify quality. Ours is instead a prolific universe of spontaneous, unmediated, non-validated information. Web 2.0 both reflects and engenders “Authority 2.0” as users, singly or in cohorts, participate in an electronic free-for-all. Platforms and formats are likewise provisional. Experimental and ephemeral expressions may evolve into dominant manifestations and forms, though extinction (think WordStar in the pedestrian realm of word

processing programs) is a real possibility as well. Libraries are on uncertain ground as they engage with this fractious, seductive, alien, and essential universe.

### Guiding Principles for Collections and Content

The trends here described suggest several general principles to guide academic libraries as they move toward the future:

1. Most information—core materials, the record of scholarship, trade publications, an increasing proportion of recorded human expression, and data—is becoming available in digital formats. The emergent electronic realm will, in time, relegate new analog materials to a diminishing subset of primary sources. Digital resources will increasingly define both the information and the scholarly landscapes. *Our future is digital: libraries must prepare for and promote this shift.*
2. Digital resources are produced, become available, and then behave differently than hardcopy objects. Among many other features, few of them can be owned in the same way as books or journals. Libraries must therefore frame their information goals in terms of providing access to content that they do not possess, as well as on-site holdings. *Libraries must broaden their focus to encompass both collections and an evolving range of content, whether owned or not.*
3. As budgets decline and priorities shift, many academic libraries will steer their acquisitions toward the basic texts and sources required for curricular support. These holdings will be heavily redundant across different institutions. Conversely, more and more noncore materials may be entirely missed. Cooperative efforts—international, national, regional, and local—can at once increase efficiencies around everyone’s need for duplicative materials and also maximize the collections reach of those libraries that are capable of pursuing scarce or unique resources. *Cooperative activities will become increasingly central to library programs and strategies.*
4. The commercialization of scholarly information, on top of long-standing trends toward monetization and privatization in the realms of mass expression and entertainment, threaten the free flow of information that the academy requires. Prohibitive costs and artificial scarcity are among the consequences. Many experiments and initiatives, with those broadly clustered under the open-access

rubric among the most promising, are now in play. *Academic libraries must actively engage in reformulating information flows and scholarly communications to protect future research and learning.*

5. Libraries have always sought to make information both accessible and usable. Catalog records link users with the sources relevant to their interests; reference (or “research and learning”) services then help those users extract the fullest possible benefit from what they have found. Digital resources—particularly large-scale, cloud-based data—require new, standards-based tools and services for description, access, use and manipulation, and preservation. *Libraries, acting independently and through external partnerships, must participate in developing all of these tools and services.*
6. Academic libraries must be aligned with and accountable to their parent institutions. Yet information is becoming more diffuse and library activities, across the board, are ever more cooperative in nature and expansive in scope. Closely consultative

processes within each campus will remain essential, but may no longer be sufficient. *Universities and libraries must devise models for governance that both ensure local accountability and encourage cooperative activities.*

The world of library collections is one in which once solid certainties no longer obtain. The range of relevant materials has shifted and grown, though the relative centrality of tangible resources under the library’s direct control is in decline. Libraries will increasingly work to identify and describe information that they will never own, and to provide the tools that enable their students and scholars to discover and use these resources effectively. The sources themselves will take on new dimensions whose continued usability will demand different kinds of support. More and more, libraries will have to engage in partnerships and collaborative efforts to achieve their goals. While the mandate to ensure ready access to a comprehensive array of information resources will remain, the “what” and “how” will seem quite different.

---

## Index to Advertisers

ALCTS.....	124
Archival Products.....	114
Library of Congress.....	cover 2
Library Technologies.....	cover 3