


# Suttenberg

## HARVARD'S LIBRARIES DEAL WITH DISRUPTIVE CHANGE.

by JONATHAN SHAW

---

HROW IT IN THE CHARLES,” one scientist recently suggested as a fitting end for Widener Library’s collection. The remark was outrageous—especially at an institution whose very name honors a gift of books—but it was pointed. Increasingly, in the scientific disciplines, information ranging from online journals to databases must be recent to be relevant, so Widener’s collection of books, its miles of stacks, can appear museum-like. Likewise, Google’s massive project to digitize all the books in the world will, by some accounts, cause research libraries to fade to irrelevance as mere warehouses for printed material. The skills that librarians have traditionally possessed seem devalued by the power of online search, and less sexy than a Google query launched from a mobile platform. “People want information ‘anytime, anyplace, anywhere,’” says Helen Shenton, the former head of collection care for the British Library who is now deputy director of the Harvard University Library. Users are changing—but so, too, are libraries. The future is clearly digital.

Yet if the *format* of the future is digital, the *content* remains data. And at its simplest, scholarship in *any discipline* is about gaining access to information and knowledge, says Peter Bol, Carswell professor of East Asian languages and civilizations. In fields such as botany or comparative zoology, researchers need historical examples of plant and animal life, so they build collections and cooperate with others who also have collections. “We can call that a museum of comparative zoology,” he says, “but it is a form of data collection.” If you study Chinese history, as Bol does, you need access to primary sources and to the record of scholarship on human history over time. You need books. But in physics or chemistry, where the research horizon is constantly advancing, much of the knowledge created in the past has very little relevance to current understanding. In that case, he says, “you want to be riding the crest of the

tidal wave of information that is coming in right now. We all want access to information, and in some cases that will involve building collections; in others, it will mean renting access to information resources that will keep us current. In some cases, these services may be provided by a library, in others by a museum or even a website.”

Meanwhile, “Who has the most scientific knowledge of large-scale organization, collection, and access to information? Librarians,” says Bol. A librarian can take a book, put it somewhere, and then guarantee to find it again. “If you’ve got 16 million items,” he points out, “that’s a very big guarantee. We ought to be leveraging that expertise to deal with this new digital environment. That’s a vision of librarians as specialists in organizing and accessing and preserving information in multiple media forms, rather than as curators of collections of books, maps, or posters.”

### **LIBRARIANS AS INFORMATION BROKERS**

BOL IS PARTICULARLY INTERESTED in the media form known as Google Book Search (GBS). The search-engine giant is systematically scanning books from libraries throughout the world in order to assemble an enormous, Internet-accessible digital library: at 12 million books, its collection is already three-quarters the size of Harvard’s. Soon it will be the largest library the world has ever known. Harvard has provided nearly a million public domain (pre-1923) books for the project; by participating, the University helped with the creation of a new tool (GBS) for locating books that is useful to people both at Harvard and around the world. And participation made the full text content of these books searchable and available to everyone in the United States for free.

GBS appeals to Bol and other scholars because it gives them quick and easy access to books that Harvard does not own (litigation over the non-public-domain works in GBS notwithstanding). For Bol, such a tool might be especially useful: Harvard ac-

# 2.0

**Isaac Kohane, director of the Countway Medical Library, sees librarians returning to a central role in medicine as curators of databases and as teachers of complex bioinformatics search techniques.**

quires only 15,000 books from China each year, but he estimates that it ought to be collecting closer to 50,000. So GBS could be a boon to scholarship.

But GBS also raises all kinds of questions. If everything eventually is available at your fingertips, what will be the role of libraries and librarians?

"Internet search engines like Google Books fundamentally challenge our understanding of where we add value to this process," says Dan Hazen, associate librarian of collection development for Harvard College. Librarians have worked hard to assemble materials of all kinds so that it is "not a random bunch of stuff, but can actually support and sustain some kind of *meaningful* inquiry," he explains. "The result was a collection that was a consciously created, carefully crafted, deliberately maintained, constrained body of material."

Internet search explodes the notion of a curated collection in which the quality of the sources has been assured. "What we're seeing now with Google Scholar and these mass digitization projects, and the Internet generally," says Hazen, "is, 'Everything's out there.' And everything has equal weight. If I do a search on Google, I can get a scholarly journal. I can get somebody's blog posting....The notion of collection that's implicit in 'the universe is at my fingertips' is diametrically opposed, really, to the notion of collection as 'consciously curated and controlled artifact.'" Even the act of reading for research is changed, he points out. Scholars poring through actual newspapers "could see how [an item] was presented on the page, and the prominence it had, and the flow of content throughout a series of articles that might have to do with the same thing—and then differentiate those from the books or other kinds of materials that talked about the same phenomenon. When you get into the *Internet* world, you tend to get a gazillion facts, mentions, snippets, and references that don't organize themselves in that same framework of prominence, and typology, and how stuff came to be, and why it was created, and what the intrinsic logic of that category of materials is. How and



whether that kind of structuring logic can apply to this wonderful chaos of information is something that we're all trying to grapple with."

How does searching digitally in a book relate to the act of reading? "There may be a single fact that's important," Hazen explains. "Is the book's overall argument something that's equally important as the single fact or is it just irrelevant? When people worry about reading books online, part of the worry is that the nuances of a well-developed argument that goes on linearly for 300 pages [are missing]. That's not the way you interact with a text online." How the flood of information from digitized books will be integrated into libraries, which have a separate and different, though not necessarily contradictory, logic remains to be seen. "For librarians, and the library, trying to straddle these two visions of what we're about is something that we're still trying to figure out."

Moreover, the prospect that, increasingly, libraries will be



Visit [harvardmag.com/extras](http://harvardmag.com/extras) for links to the libraries' digital collections and a video spoof of an earlier cultural transition





**The book, or codex, took hundreds of years to replace handwritten manuscripts, which persisted as an economical way to produce small numbers of copies into the nineteenth century, nearly 400 years after Gutenberg invented movable type. Robert Darnton, director of the University Library, shown with Diderot's *Encyclopédie*, predicts great longevity for the book.**

raw information are impossible to absorb and understand without computational aids.

Medicine has had to cope with this problem ever since nineteenth-century general practitioners found they could no longer keep up with the sheer quantity of published medical literature. Specialization eventually allowed doctors to focus only on the journals in their particular area of expertise. Throughout such transitions, libraries played an important role. Doctors, upon completing their rounds, would comb the stacks for records of similar cases that might help with diagnosis and treatment. Today, the amount of new information being generated in the biological sciences is prescribing another momentous shift that may provide a glimpse of the future in other disciplines. For a doctor, learning about a genetic test and then interrogating a database to understand the results could save a life. For libraries and librarians, the new premium on skills they have long cultivated as curators, preservers, and retrievers of collective knowledge puts them squarely on top of an information geyser in the sciences that could reshape medicine.

### **MINING THE BIBLIOME**

ISAAC KOHANE, director of the Countway Library at Harvard Medical School (HMS), recently asked a pointed question on his blog: Who is the better doctor—the one who can remember more diagnostic tests or the one “who is the

quickest and most savvy at online searching for the relevant tests?” He predicts that “we are going to be uncomfortable with some of the answers to these questions for many years to come” because success based not on bedside manner, but on competence interacting with a database, implies a potential devaluing of skills that society has honored. And who is pondering these issues most acutely? A blogging librarian and pediatric endocrinologist with a Ph.D. in computer science.

One hundred years ago, says Kohane, a report on medical education in the United States concluded that doctors were inadequately prepared to care for patients. Half the medical schools in the country closed. “I think we are at a similar inflection point,” he says. “If you look at bioinformatics and genetics, you see vivid examples—which can be generalized to other parts of medicine—where the system has inadequately educated and empowered its workers in the use of search, electronic resources, and automated knowledge management.” Genetic testing, he adds, offers a “prismatic example”: studies in the Netherlands and the United States have shown that “physicians are ordering genetic tests because pa-

stewards of vast quantities of data, a great deal from books, and some unique, raises very serious concerns about the long-term preservation of digital materials. “What worries us all,” says Nancy M. Cline, Larsen librarian of Harvard College, “is that we really haven’t tested the longevity for a lot of these digital resources.” This is a universal problem and the subject of much international attention and research. “If you walk into the book stacks,” she points out, “you can simply smell in some areas the deterioration of the paper and leather. But with something that hums away on a server, we don’t have the same potential to observe” (see “Digital Preservation: An Unsolved Problem,” page 82).

Despite these caveats, Bol’s vision of future librarians as digital-information brokers rather than stewards of physical collections is already taking shape in the scientific disciplines, where the concerns raised by Hazen are less important. In fields faced with information overload—such as biology, coping with a barrage of genomic data, and astronomy, in which an all-sky survey telescope can generate a terabyte of data in a single night—the torrents of

tients are asking them to, [even though] they don't know how to interpret the tests and are uncomfortable doing so."

Kohane sees similar problems when making the rounds with medical students, fellows, and residents: "When we run into a problematic complex patient with a clearly genetic problem from birth, and I ask what the problem might be and what tests are to be ordered, their reflex is either to search their memories for what they learned in medical school or to look at a textbook that might be relevant. They don't have what I would characterize as the 'Google reflex,' which is to go to the right databases to look things up." The students doubtless use Google elsewhere in their lives, but in medicine, he explains, "the whole idea of just-in-time learning and using these websites is not reflexive. That is highly troublesome because the time when you could keep up even with a subspecialty like pediatric neurosurgery by reading a couple of journals is long, long gone."


The journals themselves have grown in number and quantity of articles, but "the amount of data being produced and analyzed in large, curated databases," Kohane says, "exceeds by several orders of magnitude what appears in printed publications." The fact that students and doctors don't think to use this digital material is an international problem. "Even at Harvard," where "we spend millions of dollars" annually for access to the databases, "many of the medical staff, graduate students, and residents don't know

how to use...," he pauses. "Well, it's worse than that. They don't know that they exist."

But in this lamentable situation Kohane sees an opportunity for medical libraries, whose role, he believes, had faded for a while. "It is becoming so clear that medicine and medical research are an information-processing enterprise, that there's an opportunity for a library that would embrace that as a mission...to be again a center of the medical enterprise."

Kohane has sought to do just that by creating an information institute—an HMS-wide center for biomedical informatics—embedded within Countway Library. The institute offers voluntary mini-courses, invariably oversubscribed, explaining what the relevant databases are, how to plumb them, and how to analyze the data they produce. A parallel effort under his supervision seeks to "mine the bibliome"—the totality of the electronically published medical literature—by allowing researchers to track down relationships between genes and diseases in the published literature that would not be apparent when searching one reference at a time. Librarians in the institute also comb databases for contradictions, and find references to sites in the genome that can't possibly exist because the coordinates are wrong. In making sure that information is good, the library is "returning to its original mission of curation," says Kohane, "but in a genomic era and around bioinformatics." This defines a new role for librarians as database experts

## BOOKS AND BUDGETS

 **he future may be digital,** but the book is far from dead, says University Library director and Pforzheimer University Professor Robert Darnton. More books are published now than at any other time in history, and the number of new works published each year is growing fast. So, too, is the cost of individual books. Book prices have been compounding at more than twice the rate of inflation, but library budgets have grown slowly. People still prefer to read printed books, and now they also want music, poetry, videos, and online access to the Internet, to electronic journals and specialized databases. Libraries are caught in the middle of this transition. Arguably, the stakes are nowhere higher than at the largest academic library in the world. With a \$165-million budget and 1,200 employees, Harvard's 73 libraries anchor within the Greater Boston area one of the most concentrated centers of learning anywhere in the world.

Because of the University's scope and scale, the challenges facing its libraries encompass in many ways those confronting all libraries. But Harvard also faces some special challenges. More than 50 separate administrative structures govern those 73 libraries, and this complex-

ity, now being addressed by a task force charged with simplifying it, leads to redundancy, inefficiency, and a failure to capture economies of scale in negotiating with vendors (see "Libraries on the Edge," January-February, page 41). Harvard's current budgetary crisis has led to cutbacks in all areas of collecting; cash infusions and last-minute negotiations staved off the most potentially damaging cuts this year as President Drew Faust quietly transferred \$1 million to the University Libraries from her own discretionary funds last fall; the Arcadia Foundation made a large gift; and Faculty of Arts and Sciences dean Michael Smith rolled back planned cuts to the budget of the Harvard College Library.

Meanwhile, the University's evolving understanding of itself as a world institution has led to an acute need to collect from areas of the globe in which Harvard research has not been traditionally strong, but where economic and political development has led to a burst of publishing so voluminous and chaotic that it is almost impossible to keep up with the output, as in China; or where markets and communications are so underdeveloped that scarcity and imperfect information make collecting there a special chal-

lenge, as in Africa. Pressures to collect in these and other new geographical areas of study, particularly in south and south-east Asia, as well as eastern Europe, have in turn caused anxiety among professors working in other areas because—even as the University expanded its global reach, establishing new centers and programs—funding for other library endeavours was *not* included.

Faculty members working in traditional areas use spirited language when discussing what they describe as a competition for limited resources, particularly the humanists, for whom libraries are the equivalent of scientific research laboratories, points out Ann Wolpert, director of the library at MIT. To the extent that trenchant exchanges reflect scholars' passion for their subjects, well, this is Harvard: "We'd be more worried if they weren't passionate," says Larsen Librarian of Harvard College Nancy Cline. But even if the University's scholarly mandate had remained frozen, the information explosion—new kinds of materials, in novel formats, and in quantities unimaginable even a decade ago—would have led eventually to this point, where keeping up with scholarly output in all its forms is no longer manageable for any one institution alone.



and teachers, while the library becomes a place for learning about sophisticated search for specialized information.

Such skills-based teaching, learning, and data curation depend on finding individuals who are trained in medicine and also have the public-minded qualities of a librarian—rare indeed, as Kohane readily acknowledges. And even though the cost of such bioinformatics education is small relative to the millions of dollars spent on subscription fees for electronic periodicals (the price of which doubled between 2000 and 2010, says Kohane; see “Open Access,” May-June 2008, page 61 for more on the crisis in scholarly communication), the resources to provide more educational support for complex types of database search training are insufficient across the University. “That’s because we are trying to bolt on a solution to a problem that probably should be addressed foursquare within the core educational process,” he says.

There is growing awareness of the need to have an “information-processing approach to medicine baked into the core education of doctoral and medical students.” Otherwise, Kohane says, “we’re condemning them to perpetual partial ignorance.” Already, a few lectures on the topic are being introduced into the medical-school curriculum, making HMS a pioneer in this area. Discussions about bringing more of the biological/biomedical informatics agenda to the undergraduate campus are also under way.

Even in the relatively tradition-bound profession of law, digitization cuts so deeply that when Ess librarian and professor of law John G. Palfrey VII restructured the Law School library last year, he says he thought about the mission less as “How do we build the greatest collection of books in law?” and more as “How do we make information as useful as possible to our community now and over a long period of time?”

This focus on information services within a community guided both personnel decisions and collections strategies. “We scrapped the entire organizational structure,” reports Palfrey (whose digital genes can be traced back to his former position as executive director of the law school’s Berkman Center for Internet and Society). Last June 30, all the librarians handed in resignations for the jobs they had held and received new assignments. There is now a librarian who works with faculty members, teaching empirical research methods, and another who helps students and faculty conduct empirical research. The collection development group includes “a lab for hacking a library”: a member of that team is working on an idea called “Stack View” that would allow the re-creation of serendipitous browsing in a digital format. Technology “allows you to reorganize information and present it in a totally different way,” Palfrey points out.

The law library’s new collection-development policy is organized along a continuum of materials for which the library takes increasing responsibility. These range from resources in the public domain that aren’t collected, but to which the library provides access; to materials accessed under license; and all the way up to unique holdings of an historic or special nature that the library archives, preserves, and may one day digitize in order to provide online access. The fact that the library no longer buys everything published in the law has been made explicit. “It is no longer possible financially, nor is it desirable—not all of it is useful,” Palfrey says bluntly. Only a third of newly purchased books are initially bound. “We’ll put a barcode on it, put it on the shelf, and see if

people use it,” he explains. “If they do, and the book starts to wear, then we’ll send it to the bindery.”

Even though these changes may seem like cutbacks (they were in fact planned and in process before the University’s financial crisis became apparent), he believes skilled librarians are in no danger of becoming obsolete: “The role of the librarian is much greater in this digital era than it has ever been before.” Good lawyers need to be good at information processing, and Palfrey found in research for his book *Born Digital* that students today are not very good at using complex legal databases. “They try to use the same natural-language search techniques” they learned from using Google, he says, rather than thinking about research as “a series of structured queries. It’s not that we don’t need libraries or librarians,” he continues, “it’s that what we need them for is slightly different. We need them to be guides in this increasingly complex world of information and we need them to convey skills that most kids actually aren’t getting at early ages in their education. I think librarians need to get *in front* of this mob and call it a parade, to actually help shape it.”

Mary Lee Kennedy, executive director of knowledge and library services at Harvard Business School, whose very title suggests a new kind of approach, agrees with Palfrey. “The digital world of content is going to be overwhelming for librarians for a long time, just because there is so much,” she acknowledges. Therefore, librarians need to teach students not only how to search, but “how to think critically about what they have found... what they are missing... and how to judge their sources.”

Her staff offers a complete suite of information services to students and faculty members, spread across four teams. One provides content or access to it in all its manifestations; another manages and curates information relevant to the school’s activities; the third creates Web products that support teaching, research, and publication; and the fourth group is dedicated to student and faculty research and course support. Kennedy sees libraries as belonging to a partnership of shared services that support professors and students. “Faculty don’t come just to libraries [for knowledge services],” she points out. “They consult with experts in academic computing, and they participate in teaching teams to improve pedagogy. We’re all part of the same partnership and we have to figure out how to work better together.”

## **“JUST IN TIME” LIBRARIES**

ALL THIS IS NOT TO SUGGEST that the traditional role of libraries as collections where objects are stored, preserved, and retrieved on request is going away. But it is certainly changing. Two facilities—one digital, the other analog—suggest a bifurcated future. The two could not be more different, though their mandates are identical.

In Cambridge, the Digital Repository Service (DRS) is a rapidly growing, 109-terabyte online library of 14 million files representing books, daguerreotypes, maps, music, images, and manuscripts, among other things, all owned by Harvard. In a facility that also serves other parts of the University, a two-person command center monitors more than a hundred servers. Green lights indicate all is well; red flashes when environmental conditions such as temperature or humidity exceed designated parameters. In a nearby room, warm and alive with the whirr of hundreds of cooling fans, their cumulative sound resembling the roar of a giant waterfall, a handful of servers hold the library’s entire digital collection. Other

**“A man will turn over half a library to make one book,” said Samuel Johnson. Nancy Cline, Larsen librarian of Harvard College, displays a manuscript letter from the Hyde Collection of Dr. Samuel Johnson; all its Johnson letters are available online as part of the University Library’s open-collections program.**

servers are dedicated to “discovery,” the technical term for the searchable online catalog, or “delivery,” the act of providing a file to an end user.

There are at least three copies of the entire repository—one in, and two outside of, Cambridge. One of them, secured by thumbprint access, is constantly being read by machines at the disk level to ensure the integrity of the data, a process that takes a full month to complete. “Several times a year,” says Tracey Robinson, who heads the library’s office for information systems, “we detect data that have become corrupted. We engage in a constant process of refreshing and making sure that everything is readable.” Any damaged material is quickly replaced with another copy from the backup.

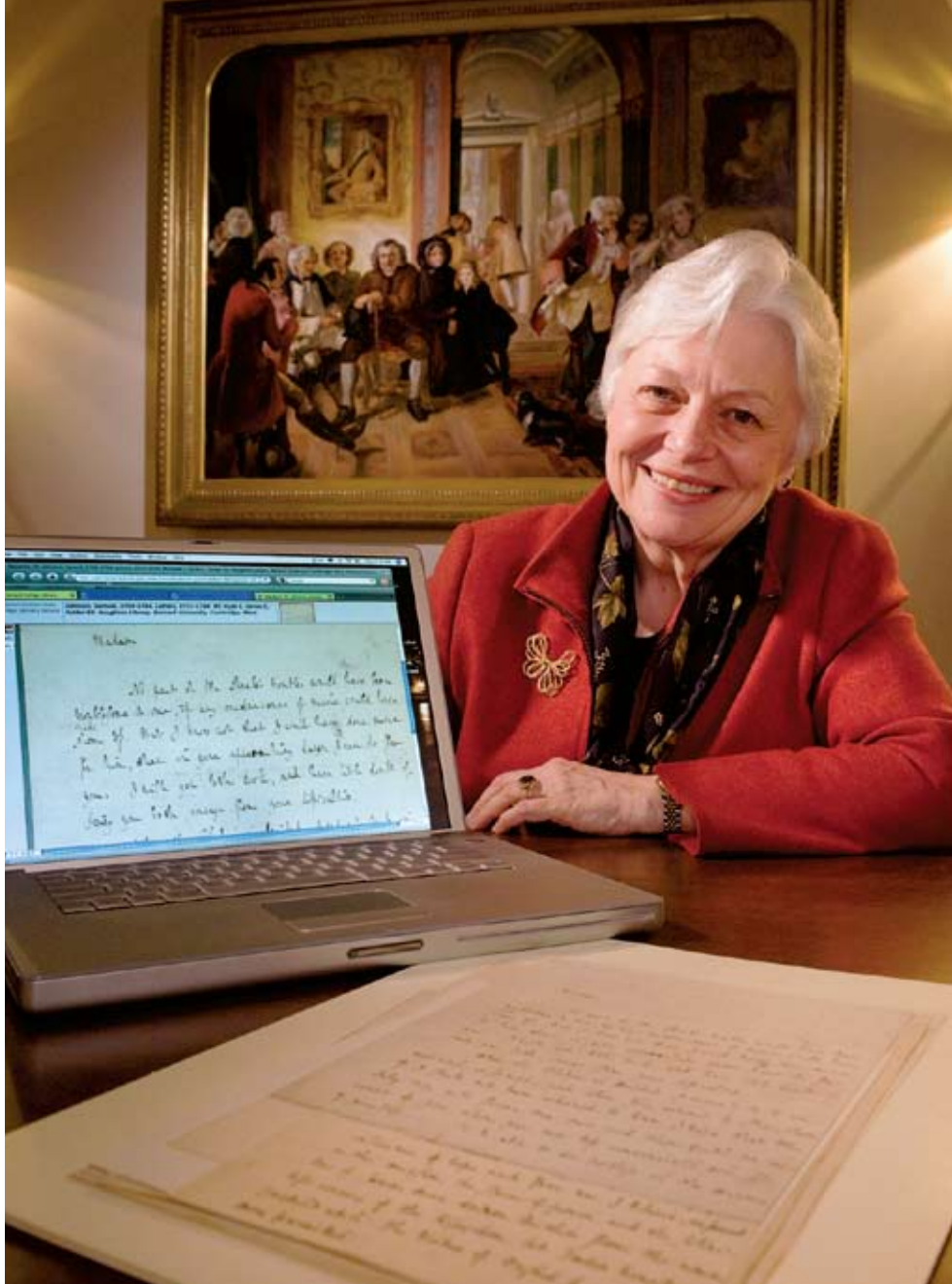
The analog counterpart to the DRS is the Harvard Depository (HD), located in the countryside about 45 minutes from Boston. A low, modular building with loading dock bays, it resembles a warehouse more than anything else. In many ways, that is precisely what it is. Just two librarians oversee 7.5 million books held in an energy-efficient, climate-controlled environment—more than twice as many as are held at Widener, which is three times as large. “The libraries based in the city are among the most expensive in terms of linear capacity,” says Nancy Cline. “The Depository as a concept is absolutely essential for us.” A number of other libraries in the Boston area, including MIT, use the HD. The facility absorbs half a million new books each year, circulates 220,000, and boasts a 100 percent retrieval rate. (In 24 years, just two books could not be found for delivery; in a typical library, one study showed, patrons find what they are looking for only 50 percent of the time.)

The secret to the HD’s extraordinary density and retrieval rate is simple: here, a book is not a book. Titles, subjects, authors—none of this so-called “metadata,” the information people typically use to find things, matters. “We know how many books we get in,” says assistant director of the University Library for the Harvard Depository Tom Schneider, who directs the facility, “but we don’t know what they are. To us, they are just barcodes. It makes our work much more efficient.” A staff of dedicated workers, who rotate through different tasks in order to break up the routine, can check in as many as 800 barcodes an hour. All the items are sorted and shelved according to size in bins that are themselves barcoded. This allows the height of the shelves to be perfectly calibrated to the height of the books; no wasted airspace. Place a request for one of the books in the HD and it will be delivered the next business day to the campus library of your choosing.

Originally, the HD was intended to store only low-circulation

items. But because the libraries of the Cambridge campus are “full to bursting,” says Pforzheimer University Professor Robert Darn-ton, the director of the Harvard University library, “doing triage” on thousands of little-used books from the shelves each year to make room for new ones proved impractical. Now, most new books are simply sent to the HD. Although some professors lament the death of shelf-browsing, others are grateful when a book they love is sent off, because they know that when next they want it, not only will it be found, it will be well-preserved: time essentially stands still for the books at the HD, where an environment set at 50 degrees and 35 percent relative humidity is expected to maintain a book in the condition in which it arrived for 244 years.

The price of such longevity and retrievability is about 30 cents per stored volume per year, which compares favorably to the cost of digital storage; expense estimates from the HathiTrust (a national group of research libraries that have created a joint repository for digital collections) for storing a digital book scanned by Google range from 15 cents for black (please turn to page 82)





and white to 40 cents for color annually. Actually delivering a physical book from the HD, on the other hand, costs \$2.15—much more than the delivery of a digital book to a computer screen.

But making comparisons between digital and analog libraries on issues of cost or use or preservation is not straightforward. If students want to read a book cover to cover, the printed copy may be deemed superior with respect to “bed, bath and beach,” John Palfrey points out. If they just want to read a few pages for class, or mine the book for scattered references to a single subject, the digital version’s searchability could be more appealing; alternatively, students can request scans of the

pages or chapter they want to read as part of a program called “scan and deliver” (in use at the HD and other Harvard libraries) and receive a link to images of the pages via e-mail within four days.

One can imagine a not too radically different future in which patronless libraries such as the DRS and the HD would hold almost everything, supplying materials on request to their on-campus counterparts. Print on demand technology (POD) would allow libraries to change their collection strategies: they could buy and print a physical copy of a book only if a user requested it. When the user was done with the book, it would be shelved. It’s a vision of “doing libraries ‘just in time’ rather than ‘just in case,’” says Palfrey. (At the Harvard Book Store on Massachusetts Avenue, a

POD machine dubbed Paige M. Gutenberg is already in use. Find something you like in Google’s database of public-domain books—perhaps one provided by Harvard—and for \$8 you can own a copy, printed and bound before your wondering eyes in minutes. Clear Plexiglas allows patrons to watch the process—hot glue, guillotine-like trimming blades, and all—until the book is ejected, like a gumball, from a chute at the bottom.)

Indeed, the HD might one day play a role as the fulcrum for “radical collaboration” with the five other law libraries in the Boston area, says Palfrey. “We’re asking, ‘Could we imagine deciding, as a group of six, that we’re actually going to buy something and put it in the Harvard Depository,’” a central location from

## DIGITAL PRESERVATION: AN UNSOLVED PROBLEM

**G**iven the convenience and potential cost saving of digital delivery for both libraries and users, combined with the power digitization offers to search within texts, why not embrace the digital future now? The issue of preservation is one of the main obstacles.

Speaking from her experience as head of collection care for the British Library, Helen Shenton explains that “the greatest risks to printed material are the environment, wear and tear, security, and custodial neglect.” Facilities such as the Harvard Depository address most of those concerns, although wear and tear is an unavoidable consequence of use. On the digital side, on the other hand, use of the data is one of the best ways of preserving it, because “‘bit rot’ is one of the biggest risks.” A book left on the shelf for a hundred years might be fine, Shenton says, but digital data must be read and checked constantly to ensure their integrity.

For digital preservationists, a prime concern is that data might be kept perfectly secure and complete, but *still* be unreadable by machines and programs in the future. A *New Yorker* cover depicting an alien, come to post-apocalyptic Earth, sitting amid the detritus of modern civilization—discarded CDs, tapes, and computers—illustrates the point: the alien is reading a book, the only thing that still “works.” “You have to think about moving the content along as technology changes,” explains Andrea Goethals, digital-preservation and repository-services manager. In order to make this feasible, librarians try to limit the number of file formats they make use of, and store detailed technical metadata with every object so that in 10 years or 100 “it can be rendered again in a usable way.” Every few years, as the programs that created a text file or a PDF become obsolete, librarians must ensure that the contents of those files remain readable by the current generation of com-



puters and software. But opening each file manually in order to save it in a current format is not feasible when there are millions of them. “Because of the enormous amount of digital material we hold, migrating content is done at scale, not one file at a time,” Goethals explains. “We have to be able to do it on a whole class of objects”—all Microsoft Word files, for example. This content management strategy *should* work, “but because digital preservation is a young science, we don’t have a lot of experience with it yet.”

Objects that begin in an analog format—a book, a recording of a poetry reading, or a piece of music—are easiest to preserve digitally because librarians can choose the optimal file format for long-term access. Material that is born digital—e-mail, for example,

which comes in many different, often proprietary, formats—is not so easy to preserve. This is proving a vexing problem for librarians charged with maintaining records of the University’s intellectual life. Harvard archivists, for example, must figure out what to do with the president’s e-mail—both a technical and legal challenge, because of privacy laws governing its handling.

For medical librarians like Isaac Kohane, the lack of case notes from the last 20 years is an equally grave concern. He recalls his excitement, shortly after becoming director of the Countway, at being able to read the notebooks of a surgeon who was about to conduct the first successful cadaveric kidney transplant, for which he later won the Nobel Prize. “You could see the foresight and the hypothesis he had, as he made that leap. So it is stunning to realize that we are in a period of unprecedented lack of documentation of academic output.” Most notes, images, and communications today are stored as e-mail. “I’m worried that, 30 years hence, people will say, ‘Well, what was happening during the great genetics revolution at Harvard University?’ They’ll have no idea.”

**Nearly half of Harvard's collection is housed at the Harvard Depository, a marvel of efficient off-campus storage. Library assistant Carl Wood reshelves books in the 30-foot-high, 200-foot-long stacks.**

which the physical book could be delivered to any institution? "It would cost us a sixth as much." Other Harvard libraries could explore the same strategy.

That doesn't mean Harvard's campus libraries would become less important. Because they are embedded in the residential academic community, they remain integral to University life. Students (and faculty members) are big users of the physical spaces in libraries, though they are using them differently than in the past.

"Libraries are not conservative places anymore," says Cline. "From the user perspective, it is an interesting time. Some people still want the quiet, elegant reading room. Others would be frustrated if they had to be quiet in every part of the buildings, in part because their work requires that they talk, that they work in collaborative teams, that they share some of their research strategies. We're rethinking the physical spaces to accommodate more of the type of learning that is expected now, the types of assignments that faculty are making, that have two or three students huddled around a computer working together, talking."

Libraries are also being used as social spaces, adds Helen Shenton, where people can "get a cup of coffee, connect to WiFi, and meet their friends" outside their living space. In terms of research, students are asking each other for information more now than in the past, when they might have asked a librarian. "The flip side," Shenton continues, "is that some places are embedding their library and information specialists within disciplines and within faculties. So I think the whole model is like one of those snow globes. You pick it up and shake it around and all the pieces will settle in a different way, which is incredibly exciting."

## A FUTURE FOR BOOKS

"A BIG MISCONCEPTION is that digital information and analog information are incompatible," says Darnton, himself an historian of the book. "On the contrary, the whole history of books and communication shows that one medium does not displace another." Manuscript publishing survived Gutenberg, continuing into the nineteenth



century. "It was often cheaper to publish a book of under a hundred copies by hiring scribes," he says, than it was to set the type and hire people to run the press. Likewise, horsepower increased in the age of railroads. "There were more horses hauling passengers in the second half of the nineteenth century than there were in the first half. And there is good evidence that now, if a book appears electronically on your computer screen, and it's available for free, it will stimulate sales of the printed version."

Jeffrey Hamburger—a scholar of an even earlier medium, the medieval manuscript—who was recently named chair of a library advisory group, says that "the notion that we are going to abandon the codex as we have known it—the traditional book—and go digital overnight is very misguided. It is going to be a much longer transition than anyone suspects, just as the transition in the past between the oral tradition of literature in antiquity and silent reading as we've known it for almost two millennia was a long transition, taking the better part of a millennium itself."

Hamburger, the Francke professor of German art and culture, has worked extensively here and in Germany on projects involving the application of new media to the study of medieval manuscripts, but he says there are "still many, many things that new media cannot do as effectively as a good old-fashioned book": for example, combining text and an associated image on opposing pages. "It's instructive how many of the words we use to describe computer interfaces—tabs, bookmarks, scrolling—are derived from our experience with the book, and that's not just because of expe-

rience or familiarity," he adds. "It's because they have a certain practicality, and all of those, it so happens, are inventions of the Middle Ages." Computers, in reverting to scrolling, have "gone back to a much older technology, which had its merits but was deficient in its own ways, which is why it was replaced."

In advocating for the continued importance of books, and raising his concern that this could become the "lost decade" for acquisitions to Harvard's library collections, Hamburger emphasizes that he is not framing the University's current crisis in terms of books versus new media. "We need both, and we'll continue to need both. I think we have to take as a premise that the library is a vast, far-flung, varied institution, as varied and diverse as the intellectual community of the University itself, working for a range of constituents almost impossible to conceive of, and it's not *just* a service organization. I would even go so far as to call it the nervous system of our corporate body."

It would be a terrible mistake, Hamburger continues, "if different factions within the faculty, be it scientists and humanists, be it Western- or non-Western-focused scholars, started squabbling over resources. As a university, we have by definition a catholic, all-embracing mission, and the question is how to coordinate resources, not compete for them. The greatness of this university in the past and in the future rests on the greatness of our library. Without the library—old, new, digital, printed—this institution wouldn't be what it is." ▢

*Jonathan Shaw '89 is managing editor of this magazine.*