Camel Drivers & Gatecrashers

Quality Control in the Digital Research Library

By Douglas Greenberg

It is said that Abdul Kassem Ismael, the scholarly Grand Vizier of Persia in the tenth century, had a library of 117,000 volumes. He was an avid reader and truly a lover of books. On his many travels as a warrior and statesman, he could not bear to part with his beloved books. Wherever he went, they were carried about by 400 camels trained to walk in alphabetical order. His camel drivers thus became librarians who could put their hands instantly on any book for which their master asked.

Until very recently, perhaps no one had devised a better scheme for the staffing and organization of a library than Abdul Kassem Ismael. Certainly no one had devised a more portable system! Of course, Ismael's system worked in part because the library had only one user making demands on the staff, enough camels to carry the collection, and enough camel drivers to care for the camels and serve the needs of the Grand Vizier. The modern library does not have these luxuries, especially in this country, where access to information is something akin to a constitutional right, and where no major library has enough staff or space to care for its collections. Access to information is at least a theoretical prop for democratic politics and social equality in the United States. Providing it is a social good of undoubted value.

Reprinted with permission from Brian L. Hawkins and Patricia Battin, eds., *The Mirage of Continuity: Reconfiguring Academic Information Resources for the 21st Century*, published by the Council on Library and Information Resources and the Association of American Universities, Washington, D.C., 1998.

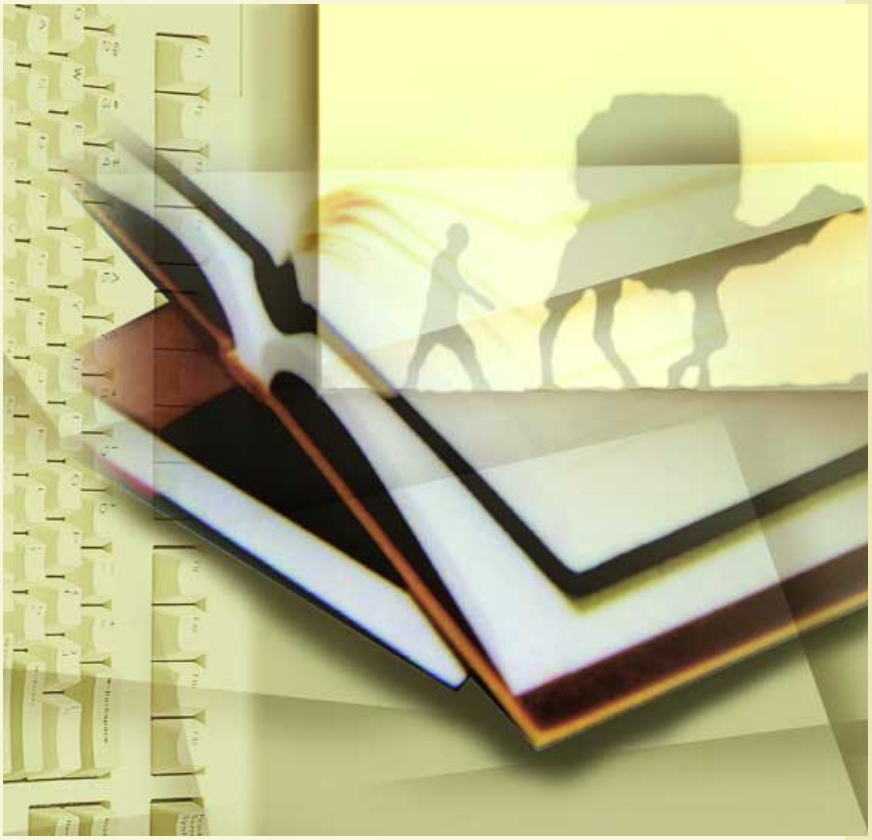

Illustration by Glenn Pierce

Because we cannot tolerate the elitist structure of Ismael's system, and because we have so much more information to care for and distribute than he did, we have begun to attempt to democratize access through electronic technology. As Ismael's library grew, he bought more camels and hired more camel driver–librarians. Until recently, we have done pretty much the same thing. But today, we hear interminably, we are about to move to the era of the digital library, a new beast of burden that will finally and fundamentally transform Ismael's system, making vastly more information available with no increase in staff. We will create the universal library, as transparently usable in Perth as in Persia. In which case it would probably be a good idea first to figure out just what a digital library actually is.

Despite the extensive, mind-numbing discussions of and work on the creation of the digital library in individual libraries, on campuses, and nationally through such efforts as the Digital Library Federation, very little has been said (or done) to specify what it signifies to use the adjective "digital" to modify the noun "library." The term "digital library" may even be an oxymoron: that is, if a library is a library, it is not digital; if a library is digital, it is not a library. We have not thought as systematically as we should about the characteristics of the print library and how and whether they can, or should be, duplicated, transformed, or abandoned in a digital world. Digital library projects abound, but they are disparate, even contradictory, in their aims, and they are also blissfully unbothered by the unintended consequences that they presage. As the wag said: "If you don't know where you are going, any road will take you there." That is precisely our situation in the transition from the print to the digital library. Because we do not know where we are going (but want to get there very quickly), the application of digital technologies sometimes becomes an end in itself.

As an exercise in nostalgia if nothing else, I propose to describe some of the characteristics of the book and the library, and then do the same thing for the electronic environment and, before then, suggesting some of the potential dangers of failing to heed the differences and some of the advantages of recognizing them.

We begin with some preliminary descriptions. The book and the library embody order, linearity, knowledge as sequence, information as a hierarchy of value, predefined relationships among disparate data, and an emphasis on the physical reality of information embodied in the printed page and three-dimensional objects. The Web page and the Internet embody disorder, circularity, knowledge as consequence, information as equality of value, relationships among data created on the fly, and an emphasis on the electronic reality of information embodied in magnetically encoded media and digital transmission mechanisms.

Thought and creativity are more frequently analogous to the Web page and the Internet: stream of consciousness, unlikely connections, imagined and created relationships all transmitted electrically across the synapse. Most of what we know, on the other hand, was originally created and organized in the linear, hierarchical, and physical world of the book and the library. We are accustomed to suppressing many of our instinctive ways of thinking, which I believe actually lack the organizational specificity of the book and the library, relying as they do on intuition, experience, and instinct.

We have learned to accommodate our impulse toward discursive thought to the inflexible categories and order we have imposed on information, but we have done so for reasons that are more logistical than anything else: we have needed, as Abdul Kassem Ismael did, a systematic mechanism for locating static print information. Our most pressing problem of the moment is that we are trying to create new pathways of access to static materials created in the culture of the book and the library, but we are using the Web and the Internet, which are inherently dynamic. As we move increasingly from digitized information created in the world of print to truly digital information, we will have to wrestle with a new problem: how to devise dynamic mechanisms for locating information that is itself dynamic as to both content and location.

The modern research library is a marvel of the human genius for organization, structure, and order, as well as for creating the tools through which that order can be understood and navigated. As complex as our libraries and library systems can sometimes be, we have only embellished Abdul Kassem Ismael's model. We begin with a question and proceed systematically from a road map to take us to a library building through its doors to an On-line Public Access Catalogue (OPAC) or other catalogue that will identify and locate the precise item or items in the library that will help us answer our question. Floor plans of the library's interior will take us to the exact shelf location where that item resides. If the item is a book, a table of contents and an index will give us two different ways to discover whether the answer is within its covers. The existence of page numbers permits us to go to a relatively small block of text and scan it for our answer. At each point along the way, we follow a route through a series of boxes within boxes until we find the tiniest box in which (we hope) our answer will be found.

Another attribute of the library, less commented upon than its hierarchical structure, is its incorporation of many mechanisms to assure the reliability of the information it contains. Just as the user proceeds to the building, to the catalogue, to the floor, to the shelf, to the book, to the page, there are entry points for printed information along the way and hurdles of trustworthiness that must be surmounted before entry is granted. Nothing gets into the library unless some reliable person makes a judgment that it belongs there. Nothing gets into the "E" section of the Library of Congress cataloging scheme unless someone makes a judgment that it belongs there. Nothing gets published in the first place unless several people make a judgment that it is worthwhile and reliable (scholarly work, for example, must

pass muster with peer reviewers and publishers). Libraries, in turn, make judgments based on experience about the reliability of publishers. Authors themselves are also gatekeepers, and they provide others with the opportunity to check their reliability through the use of (sometimes fallible) bibliographies and footnotes.[1]

In other words, the structure and physical organization of a research library guarantee that if there is information within its walls to be found on topics we can define narrowly, we will find it within a reasonable amount of time. In addition and equally important, the elaborate structure of gatekeeping in the research library, which includes peer review and many other mechanisms for the assurance of reliability, provides a reasonable guarantee that the page in the book on the shelf on the floor in the library that contains what we are looking for, has the stamp of approval of numerous referees of different kinds whose judgment can be trusted.

Moreover, there is a close relationship between the library's physical and organizational structure and its gatekeeping function. Each level of the structure contains another check on the quality of the information within. The construction of the hierarchy that the library supports is thus not only a way to provide access to information that follows a rational road map, but a way to be certain that nothing gets into the library that does not belong there. And the library is dependent upon external others (publishers and scholars primarily) to make that happen.

The elegant, hierarchical, and logical simplicity of this scheme has served humankind very well—at least since Alexandria, to say nothing of 10th century Persia. And it is likely to do so for a long time to come. The physical library will not be disappearing anytime soon. But emerging modes for the creation, storage, and transmission of knowledge and information, as well as for access to it, threaten to establish an alternative to the library that is both more and less than a "virtual duplicate" of the physical structure of the library and that embodies a very different sense of what knowledge is and how it should be organized, accessed, and communicated.

Compare the structure and reliability of the research library and the printed book to the structure and reliability of the Internet and the Web page. The Internet has no organizational hierarchy of containers of information that proceeds smoothly and sequentially downward from the vast storehouse to the single page. On the Internet, we go directly from a first point of access to the individual page. We may make a stop at a search engine or two, but by comparison to the hierarchical structure of the library, the Internet's structure is flat. It depends not at all upon the structured arrangement of knowledge.

Arguably, it subverts all structured knowledge by assigning the same significance to a page of trivia assembled by a high school student as it does to a page of data on the solar system assembled from the Hubble Telescope. Furthermore, the information on the Internet is dynamic. It is constantly changing, and it is easily changed both by its creator and by others. This is as true of data from the Hubble Telescope as it is of a high school student's Web page. New data will supercede the old on the Internet. This is part of its power. But the text of the first printing of *Moby Dick* in a rare book library will never change. Indeed, its location—if we mean by location its place in a sequence of other books on a shelf—will never change either. Print information is as static and as stable as the medium on which it appears (provided it is acid-free!). Electronic information can be as ephemeral, as changeable, and as dynamic as the media on which *it* is stored.

If the key to the library's power is its rigid, counterintuitive arrangement of static information in a comprehensible and hierarchical structure, the key to the Internet's power is its flexible arrangement of dynamic information that permits the human mind literally to jump from one thing to another and back again with no more than stream of consciousness as a guide. It is anybody's guess which of these systems is better adapted to human creativity and curiosity.

What is certain, however, is that the connection in the library between its structure and the reliability of the information it contains is not duplicated on the Internet. Indeed, just as the library and Internet are opposites on the matter of hierarchy and organization, so too are they opposites on the matter of gatekeeping and reliability. None of the mechanisms that assure the reliability of the information we find on a page in a university press book in the "E" section of my local research library pertain to information I find on the Internet.

How did the information get into the library? Through multiple gatekeepers right down to the page level. How does it get onto the Internet? Without any gatekeeping whatsoever and with no time lag between creation, publication, and access, all of which are separate processes of quality control in the print world. The existence of the multiple gates and gatekeepers of the print world thus makes the structure of information in the library possible, and the structure permits and empowers the organization of the gatekeeping function. On the Internet, neither structure nor gatekeeping exist.

This relationship between gatekeeping and organizational structure has a direct consequence for access in both environments. Although we have been ever more clever about finding aids and access tools in the print library, the organizational and gatekeeping functionaries of the print world slow our access to information. We need to consult road maps, campus maps, OPACS, library floor plans, shelves, tables of contents and indexes, and individual pages. We may even, *in extremis*, have to speak with a librarian or two. And we must do these things in sequence; removing a random book from a random shelf on a random floor of a random library and consulting a random page is not likely to help us much. On the Internet, we can move more quickly, and sequence is meaningless. The increased speed is thrilling, but since neither we nor the information we are pursuing has passed through any tests of trustworthiness and reliability, we may wind up with information that we discover to be useless or, worse, information we believe to

be trustworthy that is actually not information at all but actually the product of fabrication or imagination.

The gatekeeping function of the print world is vital to the entire scholarly enterprise. All the participants in the scholarly community both depend upon it and exercise it. Publishers are gatekeepers when they choose to publish some work and not others. They benefit from gatekeeping when they identify peer reviewers whose *bona fides* has been established by others. Libraries, in turn, benefit from the work of the publishers and establish their own standards simply to keep their customers. We would be very skeptical about libraries if we thought that they put anything on their shelves that happened to be dropped on their doorsteps. Scholars require effective gatekeeping from publishers and libraries in order to do their work, but, in their work as peer reviewers of manuscripts and in outside tenure and promotion reviews, they also serve as gatekeepers.

Universities similarly benefit from the work of the publishers, libraries, and scholars (and, significantly, they frequently fund the whole process through library budgets, university press subventions, and faculty salaries and research support). What is more, they participate crucially as gatekeepers themselves when they hire and fire faculty. We trust the best universities to hire faculty whose work can be relied upon. Indeed, that is in some sense the most salient characteristic of a great university: its faculty produce scholarship that is utterly credible from the collections of libraries whose integrity is flawless.

The differences between the library and the Internet are made even more complicated by the fact that, as noted earlier, there are really two kinds of material going up on the Net these days: digital information that exists only in that form and digit*ized* information that originated in the world of print. The reliability of both is suspect in the digital world because their provenance is difficult to trace and easy to fabricate. Moreover, the most trustworthy material produced in the world of print is ironically the *least* likely to find its way into digital form. Copyright restrictions and the desire of rights holders to protect their intellectual property have thus far caused many holders of rights to existing peer-reviewed information (scholarly journals are only one example) to hesitate about providing digital access to their material.

Nature abhors a vacuum, and digital material created for the Net or digitized material that is not peer reviewed is far more likely to appear on computer monitors these days than scholarly material of the highest quality and reliability.[2] Students, never inclined to be concerned about these sorts of issues in any event, are likely to think that the boundary between the Net and the Library is transparent or nonexistent. Indiscriminate use of unsubstantiated data—and the lack of quality gatekeeping on the Net to distinguish it from reliable data—can threaten the very standards of scholarship and meticulousness that are at the core of the modern humanities and social sciences. At the very least, the quality of undergraduate education in the arts and sciences is threatened. The social catastrophes that a paranoid might predict as a consequence of all this are probably best left undiscussed here.

Of course, the undependability of electronic information is not the only difference between the Internet and the library. True digital information will exist in new forms and formats and combinations of forms and formats that do not exist at all in the material world of the library and the book. We have not yet learned how knowledge created in these new forms and formats will be used, much less what sorts of retrieval mechanisms we will have to create for them.

The emergence of genuine multimedia resources, including fully searchable audio and video archives, combined with the possibility of seamlessly integrating images, sound, and motion with text, means not only that the digital library will be very different than the traditional library, but also that the scholarly work we create from research conducted "in" the digital library will be characterized by a multiplicity of media and formats. New kinds of information resources will, it is not too much to say, dictate the creation of new forms of knowledge and new ways of communicating it.[3] We are only at the beginning of knowing what those new forms of scholarship will be like and what the tools will be to help people find and use that information and knowledge.

The implications of all this—not only for research collections

The gatekeeping function of the print world is vital to the entire scholarly enterprise. All the participants in the scholarly community both depend upon it and exercise it.

The new technologies permit us to digitize and make accessible to undergraduates powerfully educational collections of primary sources without endangering them.

as collections and for research as the acquisition of new knowledge—are profound. The very meaning of what a research collection is must change almost unrecognizably to accommodate new forms of research materials, new mechanisms of organization, new kinds of access points and finding aids, and, subsequently, new forms of scholarly communication.

Meanwhile, we must continue to pursue the reservation of the print record and electronic access to it. We must also urgently consider the preservation questions of the digital age. The recent film produced by the Commission on Preservation and Access, *Into the Future*, treats some of these questions superficially, but the implications of digital technology for current standards, policies, and procedures of preservation and conservation have not attracted truly thoughtful consideration in either the library or archival communities. Instead of focusing merely on the brief half-life of electronic information and the frailty of magnetic media, we also ought to be framing policies and procedures for ensuring that electronic materials are retained in pristine and original condition and are not changed by "gatecrashers" from the Internet. The "indeterminacy of the text" to which literary theorists point in the print world, is, in the digital environment an almost overwhelming reality.

Just as we still do not allow pens into the rare book room

of the library, we must assure ourselves that we have done all we can to keep the electronic graffiti artists and doodlers away from the electronic primary resources and digital scholarship. This will be easier said than done since we will not want to defeat one of the Internet's most alluring characteristics: its capacity to provide unfettered access to basic knowledge for anyone who needs it. Firewall technologies, digital signatures, and the like have begun to provide this sort of security in the commercial world. Digital libraries and digital librarians will need to apply the same sorts of technologies with a ruthlessness equal to that of commercial enterprises.

If we address these problems, they are soluble. In the meantime, the digital library marches on, with or without attention to these caveats. And it does contain within it some truly transformative possibilities that do not threaten the liberal arts tradition at all. Indeed, they promise to enliven and reinforce that tradition.

A less threatening, even promising, aspect of technologically based research collections involves the blurring of distinctions that are embedded in our system of scholarly and educational work but are more logistical than logical. We have tended, for example, to distinguish between library materials that are used for research and those that are used for instruction. Large universities that own significant collections of primary resources frequently separate them from the teaching collections of the library; they may even have undergraduate libraries with a core collection of secondary and printed primary materials. The real stuff of scholarly research is invaluable and fragile and, quite sensibly, protected from the sometimes overeager hands and highlighters of undergraduates.

This is a perfectly reasonable way to protect research collections, but no one should ever have believed that such protection served an educational purpose. Anyone who has taught undergraduates knows, in fact, that primary sources are the very best way to engage them in the scholarly enterprise. The new technologies permit us to digitize and make accessible to undergraduates powerfully educational collections of primary sources without endangering them (either the undergraduates or the sources). And, since such digitization can be undertaken by colleges and universities themselves, their authenticity can be assured (if it can also be protected).

The tendency of the Internet and analogous technologies to substitute unstructured information for structured information is similarly a potential benefit to both scholarly research and undergraduate education. The capacity to undertake full text and key word searches of vast bodies of information on the Internet is a powerful tool for research, as is the ability to "click" from place to place on the Internet, without returning repeatedly to catalogues, shelving schemes, library floor plans, and book stacks. Moreover, the imminent ubiquity of images, video, and audio that are also

searchable is thrilling from an instructional and research perspective.[4] The potential dangers of having students and colleagues using materials that have not been properly vetted and authenticated should not prevent us from exploiting these technologies, provided that we do so with care and a due regard for the traditions of the academy.

Another aspect of the same blurring of distinctions has to do with the creation of databases or Websites of selections of primary source materials rather than making the materials themselves completely available electronically. This is not only a matter of reducing the expense involved in digitizing a complete archival collection; it is also a way to impose intellectual structure in the electronic environment. Databases and well-crafted Websites, like libraries and good scholarly books, actually work against the natural tendency of the Net to flatten information and remove hierarchies of meaning. They simultaneously can permit research of a very high order and educational experiences that are as valuable for elementary and junior high school students as for undergraduates and senior citizens.[5]

The Internet thus promises to enrich at least as much as it promises to threaten traditional academic culture. Nonetheless, the nomenclature to which we are accustomed, itself an artifact of the world of print, is likely to need revision, even replacement. The distinction between education and research, as we have seen, is already blurring. Similarly, distinctions between collections and the scholarship they support, between collections and exhibitions, between scholarship and exhibitions, between documents and artifacts, between texts and objects are all likely to get very murky indeed as technology allows us to reformat, reorganize, and redefine the materials of our cultural life.

Imagine, for example, side-by-side on your screen, the score of a great work of music and a performance of it. Or a transcript of a great political speech juxtaposed to a video recording of its delivery and a newspaper account. Or a 360° view of a great piece of sculpture accompanied by a recording of a great actor reading from the sculptor's diary, and a scan of the diary itself in a third window. Remarkably, none of these are quite the technological wonder that they would have been just a few years ago: the technology exists, and the mechanisms of access are becoming increasingly inexpensive and available.

One can even imagine that a new, digital form of scholarship will emerge very much sooner than we expect. If our resources are no longer confined to the printed page, why should our scholarship be? Indeed, digital scholarship derived from digitized or digital resources should command the creation of a new kind of digital footnote that provides hotlinks to the sources in the same way that print footnotes do. One might even argue that such "hypernotes" will provide a more effective method of authenticating scholarly citations than print since it will be possible to follow links to the notes without leaving one's desk. A search of the original sources for a piece of multimedia scholarship will immediately and easily provide verification of the quality of the author's research—more immediately and more easily than do print footnotes (which are only occasionally checked by readers). Provided that we can build sufficient safeguards of authenticity into our systems of distribution and scholarly communication, therefore, the promise of the Internet to create not only digital libraries but digital scholarship and digital classrooms is very powerful indeed.

In the end, it is all a matter of quality control. Both Abdul Kassem Ismael's camel drivers and our gatekeepers safeguard the integrity and reliability of the libraries they protect in order to assure users that they can count upon the information they find there. The digital library thus far lacks its full complement of camel drivers and gatekeepers. Our friend the Grand Vizier Abdul Kassem Ismael would surely be astounded—both by what the digital library threatens and by what it promises—astounded, that is, if, in fact, he ever existed. I discovered his story on the Internet.

Notes

1. See Anthony Grafton, *The Footnote: A Curious History* (Cambridge, Mass.: Harvard University Press, 1997).
2. Projects like Journal Storage Project (JSTOR) and others are, of course, encouraging exceptions to this generalization.
3. At earlier stages in the development of these new technologies I tried to explore some of the implications of this sort of scholarship. See "Get Out of the Way If You Can't Lend a Hand: The Changing Character of Scholarship, Technology, and the Significance of Special Collections," Sul H. Lee, ed., *The Role and Future of Special Collections in Research Libraries* (New York: The Haworth Press, 1993), 83–98; *The Journal of Library Administration* 1 (1993): 83–98; *Biblion* (Fall, 1993): 5–18 [published simultaneously], "Technology and Its Discontents: Some Problems and Possibilities for the Humanist in the Virtual University," Proceedings of the Conference on Changes in Scholarly Communication Patterns: Australia and the Electronic Library (Consultative Committee of the Australian Academies of Humanities, Science, Social Sciences, and Technological Sciences and Engineering, Canberra, Australia, 1993): 131–146, and "Return to the Valley of the Dolls: Reflections on Changing Lanes Along the Information Superhighway," *Biblion* (Autumn, 1996): 3–16. An ironic sidelight to this series of publications was that they involved a conceit about what the shape of scholarly work on the novelist, Jacqueline Susann, would be like when true multimedia research became possible. In a case of life imitating art, Peter Graham of the Rutgers University Library recently sent me an e-mail from the SHARP-L listserv, run by the Society for the History of Authorship. It reads, in part: "I am working on an ever expanding project that has brought me to a place where I need to do some serious work on Metalious' *Peyton Place*." I cannot foreswear suggesting of Grace Metalious as Truman Capote did famously of Jacqueline Susann that the notion of Metalious as a serious writer is an oxymoron. See also Douglas Greenberg, "Technophobia, Papyrophylia, and the Real Thing: Psychoneurotic Barriers to Technological Innovation in Cultural Institutions," *Association of Computing Machinery Newsletter*, Special Interest Group on University Computing Support (February 1997).
4. An early example is the Oyez Oyez Oyez Website (http://court.it-services.nwu.edu/oyez/) at Northwestern University which permits access to oral arguments of the Supreme Court over the Internet. The Chicago Historical Society plans a comparable site that will include 9,000 hours of interviews conducted by Studs Terkel on WFMT in Chicago and, when the technology permits, 4 million feet of WGN news film. The Historical Society is also considering what mechanisms it will need to employ to protect its intellectual property rights, assure scholarly and educational access, and establish the authenticity of the material.
5. The superb work of Professor Edward Ayres and his colleagues at the University of Virginia in the Valley of the Shadow Project (http://jefferson.village.virginia.edu/vshadow2/) exemplifies this sort of work, as does the collaboration between Northwestern University, Professor Carl Smith of Northwestern, and the Chicago Historical Society, "The Great Chicago Fire and the Web of Memory" (http://www.chicagohistory.org/fire/).

Douglas Greenberg is President and Director of the Chicago Historical Society.