BY CLIFFORD LYNCH

FROM AUTOMATION TO TRANSFORMATION

FORTY YEARS OF LIBRARIES AND INFORMATION TECHNOLOGY IN HIGHER EDUCATION

This article examines the ways in which information technology developments have changed the academic library over the last few decades, and speculates about further changes to come. In an effort to expose the major themes, I have glossed over many important details, and I hope that those knowledgeable about these details will forgive any hints of revisionist history. Things are never as simple or tidy as this kind of brief overview implies.

Richard West and Peter Lyman have suggested a three-phase procession of the effects of information technology on organizations: modernization (doing what you are already doing, though more efficiently); innovation (experimenting with new capabilities that the technology makes possible); and transformation (fundamentally altering the nature of the organization through these capabilities). This is a very helpful way of understanding what has happened to academic libraries in the latter part of the twentieth century, but one needs to recognize that libraries function within a much broader context that includes the publishing and information marketplace, changing modalities of scholarly communication, and evolving capabilities in the user community. As other articles in this issue illustrate, information technology has profoundly changed all aspects of higher education and scholarship, and these changes continue to unfold today. Innovation and transformation for academic libraries take place within this broader context; libraries cannot be considered in isolation from this context.

The first part of the story is dominated by the theme of automation (modernization): libraries applied a growing range of information technologies to the management of collections of primarily print information. This was a supremely

Clifford Lynch is the Executive Director of the Coalition for Networked Information (CNI), jointly sponsored by EDUCAUSE and the Association of Research Libraries.

Photo by Ed Castle/Folio Inc.

rational period characterized by the primacy of the systems analysis perspective—careful studies of cost/benefit tradeoffs in the introduction of technology to modernize library operations. Starting in the late 1980s or early 1990s, academic libraries were confronted with environmental changes driven by information technology, which quickly moved the focus of attention away from automation toward a series of much more fundamental questions about library roles and missions in the digital age. Libraries were forced to *react* to developments in information technology (and their cultural and economic consequences) rather than methodically exploiting them. The emergence of the World Wide Web in the mid-1990s is perhaps the great symbol of this shift, with all of its implications for scholarly communication; but there is much more: the rise of computational science, the new role of databases in all areas of scholarship. At the start of the new century, libraries are struggling to absorb innovation and to recognize the implications and meanings of transformation.

The First Automation Age: Computerizing Library Operations

There is a rich and fascinating early history of information technology in libraries, reaching back to the 1950s and early 1960s, as part of the post-*Sputnik* revolution in science and technology.[1] Yet for most academic libraries, this technology first arrived in force in the late 1960s or early 1970s in the form of locally developed or commercial products intended to automate library processes. Minicomputers were introduced to automate circulation, and books were barcoded. Computer-based ordering systems were introduced to pass orders to book and serials jobbers. These changes simply made existing manual processes more efficient and helped to control their costs.

This was a period of significant management challenge for libraries. Many of the companies offering products were small; some used custom hardware and software, and a number of these companies failed. Some libraries developed their own systems rather than purchasing commercial products; yet few libraries had the expertise to manage large, operationally critical software projects. Libraries also learned some hard lessons about system life-cycle management. For example, the conversion from a manual circulation system to the first automated system was far easier than the conversion from the first automated circulation system to the second one, when some libraries discovered that there was no way to get information out of the old system and into the new one and often had to run the two systems in parallel for a year or more.

Perhaps the greatest achievement of this period, which continued until the early 1980s, was the development of shared copy-cataloging systems. These systems established very important precedents in the use of computers and computer networking for collaboration and cooperation within the library community and paved the way for other key developments that would change libraries in the 1980s and early 1990s. They also saved libraries a huge amount of money.

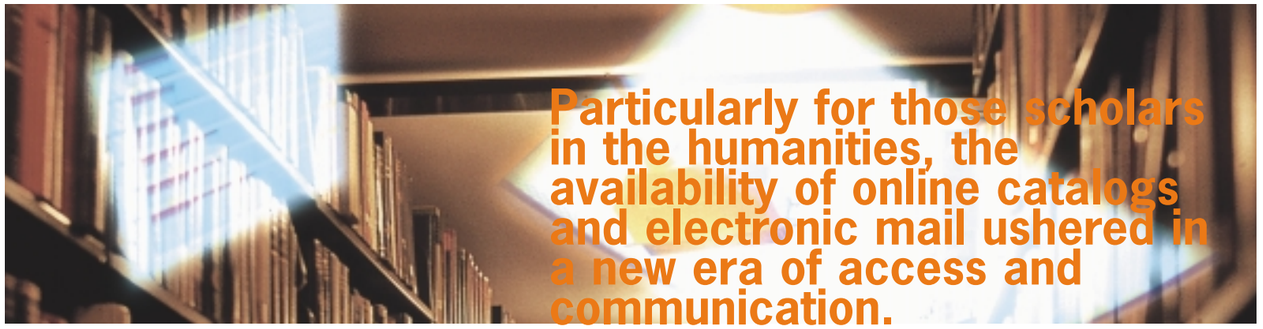
The key insight behind shared cataloging was that because most books acquired by any given research library were also acquired by other libraries, there was no reason for each library to expend the expert labor to independently catalog the book and prepare cards for its card catalog. Instead, groups of libraries used large, centralized databases operated by external organizations. On receiving a new book, a library would consult the database, and if the book was already cataloged, the library would simply attach its name to the list of libraries holding the book. The cataloger would then push a button on a terminal, and the cards would be printed for filing in the library's local card catalog (and an electronic copy of the bibliographic record would be written to an archive tape for the library). If the book was not already in the database, the library would create a cataloging record, which other libraries could then use.

Shared cataloging was pioneered by a number of library consortia in the 1960s and 1970s. Today, these efforts have consolidated into two major shared cataloging systems, one operated by OCLC in Columbus, Ohio, and the other by the Research Libraries Information Network in Palo Alto, California.

The Second Automation Age: The Rise of Public Access

By the 1980s, the shared cataloging databases had become quite large as a result of retrospective conversion programs for older books and some years of use in cataloging new acquisitions. The central databases began to reflect the collective holdings of the major research libraries. In addition, individual libraries now had machine-readable bibliographic records for significant percentages of their holdings. Also, by this time the cost of information technology—and in particular, interactive systems—had dropped to the point that it was possible to think about much larger scale applications of information technology. These developments laid the foundation for a second round of automation activities, which would start to radically change the library services visible to the patron.

The first result of this next round of changes was the online public-access library catalog as a replacement for the traditional card catalog. The physical card-catalog drawers were hauled away and replaced by banks of computer terminals. The online catalog was a tremendously powerful tool that opened up the massive collections of the research libraries to the patron community, who could now, in seconds, conduct searches that had been literally impossible to conduct in the old card catalogs. Through the online catalog, coupled with the growth of campus networks, one could search the library's holdings at any time, and could do so remotely, rather than having to go to the library. (But the user could find out only what books the library held and could not actually view the works on screen.) In addition, consortia developed union catalogs that merged materials from multiple libraries, promoting resource sharing and the growth of library consortia; a group of libraries that wanted to work together could create a "virtual" combined collection. As the Internet began to grow, library catalogs were connected to the Net so that they could be consulted remotely from anywhere in the world. Particularly for those scholars in the humanities, the availability of online catalogs and electronic mail ushered in a new era of access and communication.



The online catalog was a huge advance. But it was almost completely irrelevant to many library users. Traditionally, library catalogs have contained entries for books and for serials but have not described individual articles in a journal. Given that journals, rather than monographs, are the key literature in many disciplines, particularly in the sciences, and that by the mid-1980s a typical research library was spending more than half of its acquisitions budget on journals, the library catalog was unresponsive to the needs of many library patrons—particularly in the sciences.

Abstracting and indexing services, such as *Index Medicus* (now *MEDLINE*) for the health sciences literature, abstracted articles in journals and supplemented the local library catalog. Since the 1940s (or earlier), libraries had been purchasing these as voluminous series of printed volumes, which were very hard to use. During the 1960s and 1970s these abstracting and indexing services began to create databases in order to produce their printed products; starting in the late 1960s, these databases were also made available through commercial online services like Dialog or BRS, designed for use by specially trained searchers. However, these online services were enormously expensive and were open mostly to researchers in industry. Although a faculty member at an academic institution might be offered a few mediated searches a year, the notion of offering searches on a broad scale to students was economically unthinkable. Not until the late 1980s and early 1990s were these abstracting and indexing databases mounted for interactive public access, both by research libraries and by new commercial services that marketed to the library community, opening up the journal literature to the library patron in the same way that the online catalog opened up the monographic literature.

Many of these databases were (and continue to be) quite costly, and most research libraries spend hundreds of thousands of dollars a year to provide their patrons with access to these tools. The databases are typically acquired on a site basis: an institution pays a flat fee for unlimited use. And unlike books or journals, which a library simply buys, the databases are not purchased but rather are *licensed*; each institution negotiates with each supplier a contract that determines who may use the database and how they can use it. Negotiation of these contracts requires significant effort. Traditional library practices such as interlibrary loan are not usually relevant to these databases. Indeed, for libraries that are open to the public (such as at state-supported institutions), special provisions have been negotiated to permit the use of these databases by people other than members of the local academic community who are "incidentally physically present in the library facilities" (as often stated by the typical license).

It is easy—from this distance in time, now that these kinds of systems are simply taken for granted—to underestimate the extent of their impact on the library users of the 1980s. They made library materials enormously more accessible and allowed the materials to be searched in entirely new ways.[2] They made the idea of "anytime, anywhere" remote access to library resources real and created tremendous demand for libraries to move beyond the online bibliographic services to the actual delivery of content online. It is also useful to recognize the role that libraries played in introducing many people on campuses in the 1980s to information systems. Other than the ATM (automatic teller machine), which began to be widely used at about the same time, the online cata-

log was the first interaction that many faculty and students, particularly outside the sciences and engineering, had with technology for information access. And it is important to underscore the change in perspective that libraries went through during this period: from deploying "closed," highly optimized systems that were designed to be used in the library (which incorporated specially customized terminals, for example) to delivering open network services that were designed to be used with general-purpose workstations and that were based on industry-standard networking protocols. This difficult transition laid the foundation for many of the developments that followed.

The 1980s and early 1990s also saw major investments in resource sharing. Union catalogs were one example; another was the development of computer-assisted interlibrary loan systems that built on the shared national union catalog databases. A library that needed a book could find out which other libraries had it and could then generate and manage a request to borrow it from one of those libraries through interlibrary loan. Fax technology was applied for the delivery of journal articles on an expedited basis. However, as interlibrary loan changed in character from a way to obtain very specialized material on an occasional basis to something that more closely resembled shared collections, copyright constraints emerged as a substantial limiting factor preventing the sharing of higher-use materials between institutions on a regular basis. Indeed, as networks have become ever more capable and ubiquitous and as more and more material is becoming electronic, it has become very clear that although a tremendous amount of resource sharing is *technically* possible, much of this will be prohibited by copyright law or by license agreement. Through resource sharing and collective action, the networked environment might at first seem to offer a way for libraries to overcome the growing costs of acquiring material, but the reality has been much more limited. The greatest financial successes have been achieved by creating collective purchasing consortia that can negotiate prices for all members of the consortium, rather than by sharing already purchased resources.

The Third Automation Age: Print Content Goes Electronic

We are now in the third—and probably final—epoch of automation. Modernization has largely run its course, and new issues related to innovation and transformation are becoming dominant.

Online catalogs, though wildly popular, rapidly created demand for actual content in digital form. Once library users had begun to enjoy the freedom of remote, twenty-four-hour-a-day access, they quickly grew frustrated with searches that ended with the identification of print material that they had to wait to get or that they could not get easily (e.g., if they were searching a catalog halfway across the world, at another institution). By the late 1980s and early 1990s, the costs of storage and bitmapped display technology had come down considerably, and networks had gotten faster. It was possible to deliver content, either as page images (bitmaps and later formats such as Adobe PDF) or as ASCII text (for materials that did not need charts, graphics, equations, or special characters). With the emergence of the Web, HTML offered another alternative. Publishers and aggregators (companies that obtained material from multiple publishers and repackaged it into a "one-stop" database) began to offer this material to libraries. The convenience of electronic content was so compelling that many users—particularly students in a hurry—rapidly began to ignore materials available only in print in favor of this convenience, even though the print material might be more appropriate for some purposes. This trend has continued to the point there is now evidence that users are querying Web-based search engines as their first portal to information; because these services index full-text Web pages, one can assume that all content found will be accessible online.

Numerous troublesome issues arose; most of these are still unresolved. Libraries had already encountered the high cost and complexity of negotiating license agreements for abstracting and indexing databases. But whereas a large research library might have fifty such databases, it might acquire books, journals, and other materials from *thousands* of publishers. The idea of negotiating and managing thousands of licenses is problematic. Even though aggregators could simplify the licensing negotiations, there were delays in getting current material from the publishers to the aggregators, and some publishers refused to deal with aggregators.

Fundamental conflicts about economics also arose: libraries wanted to see price reductions while publishers, who had to invest in converting their products to electronic form, wanted to amortize this investment and sometimes to charge more for the electronic versions; publishers also argued that the electronic versions would be more useful and more heavily used and thus merited higher prices. In addition, there were questions about the relationship between electronic and print versions. Was the electronic version a supplement to the printed material or a replacement for it? In the former case, how much would the library pay for a supplement? In the latter case, did the electronic version contain everything in the print version? What was the library really getting? Having access to a publisher's online service for the duration of a license agreement was very different from taking possession of printed journals that could be kept in perpetuity. These questions reached to the conceptual foundations of library collections and the policies and values that guide their development. Clearly, the shift to electronic content has now gone beyond the automation of existing library services and activities.

Another problem was integration. Libraries want to provide a coherent service for their users; they want their users to be able to move smoothly from abstracting and indexing databases to full text and from citation of a work to the full text of that work. Library users want to move seamlessly across a range of products and services offered by different suppliers. Publishers want to maintain a brand identity and keep independent servers housing the works that they publish. There is a great deal of consistency and coherence to printed works from a variety of publishers that are shelved together and unified by an abstracting and indexing database. It is unclear whether libraries will simply license access to an assortment of publishers' Web sites or whether they will take a more extensive integrative role, which will demand a much greater investment. These problems are still very real, though efforts at cooperation and standards development are beginning to have some effect.

Finally, it should be noted that virtually all of the print content that has moved electronic has been journals rather than books (with the exception of works like encyclopedias or dictionaries, which are read randomly, in small units). There are many reasons for this, but perhaps most important is that computer screens are not an attractive reading environment for long texts such as monographs. It is reasonable to skim a journal article online or to print it on an inexpensive printer. Digitized versions of books are much more cumbersome and so far have seen limited use. As technology continues to improve, however, this is beginning to change.

One other development from the third age of automation should also be mentioned. Many libraries not only purchase published works but also maintain "special collections"—manuscripts, photographs, maps, and other unique works that they often own as intellectual property rather than just as a copy of a mass-marketed, published artifact or that are old enough to be in the public domain and thus unconstrained by copyright. Libraries—along with museums and archives—are beginning to digitize these materials and to make them available to the public and to other libraries. Ironically, these specialized materials—which historically have been nearly invisible, the province of a few privileged scholars and of occasional exhibits of the "treasures of the library"—are now among the most visible content offered by research libraries. Because there are no license issues for much of this material (it is either out of copyright or owned by the library), it is made publicly available on the Web.[3] These materials offer a tremendous resource not only for scholarship but also for teaching, since students have access to source materials on a very broad basis.

The Networked Information Revolution: Innovation and Transformation

In the late 1980s, the world of scholarly communication, teaching, and research began to change as a result of networking and advanced information technology. We entered a decade characterized by an enormous, exhilarating flowering of innovation, creativity, and experimentation. The idea of networked information emerged: a vast constellation of digital content and services that were accessible through the network at any time, from any place, could be used and reused, navigated and integrated, and tailored to the needs and objectives of each user. This was embodied in the idea of the "scholar's workstation"—a learning, analysis, and authoring hardware and software system that mediated between the user and the array of available networked information resources, identifying, negotiating with, and federating resources as needed. "Resource discovery"—the identification of appropriate resources through directories, metadata, or other means—raised questions about how responsibility for organizing network resources should be assigned: centrally, institutionally, or on a distributed basis in which each resource became self-describing.

Networked information implied a breakdown of geography as an organizing principle. All resources on the network were equally close, and they could complement or compete with each other; relationships between information providers and information users became much

more complex. International information sharing and collaboration were greatly facilitated.

The use of the Net became critical in many forms of scholarly communication. Preprints and technical reports became widely distributed on the Net, democratizing access to this critical information and speeding up the rate of communication. Monographs and reference works morphed into databases and scholarly Web sites. Government databases and datasets became critical resources in the physical and life sciences, the social sciences, and even some areas of the humanities. Access to geospatial and remote-sensing data transformed the earth sciences. Community databases became essential parts of the scientific discourse in areas such as molecular biology and genetics. Scholarly communication became much more interactive through the use of technologies as mundane as mailing lists or as sophisticated as collaboratories.

In the early 1990s, the idea of the "digital library" was popularized, in part because of a large-scale ARPA/NSF/NASA grant program. Although the definition of the digital library continues to be debated,[4] work under this grant program produced a number of very sophisticated information systems targeted at specific scholarly communities. Most of these, however, were not directly connected to the efforts of the academic library community and were designed to explore technology rather than to offer sustainable services. While they developed technologies that will be important to libraries as institutions, they were not intended to create models of the future of such institutions. Indeed, it appears that academic libraries will not become digital libraries but rather will acquire access, on behalf of their users, to the ever-growing digital collections, including those of the "digital libraries" that will be developed by scholarly or commercial organizations.

Multimedia became a routine part of content and communication for learning and research: video, images, simulations, virtual reality walkthroughs, and audio are all carried by the Web. Instructional technology gave rise to digital "learning objects" that could be used in classroom settings or for independent learning. Digital content also facilitated the creation of virtual "reserve rooms" and was harnessed to support distance education and asynchronous learning.

Expectations about services changed at this time. Capabilities such as personal views of collections of information resources, current awareness and change tracking systems, and reccomender or collaborative filtering systems were developed on the Web for consumer use. Some of these also began to appear in library service offerings, though others, notably reccomender systems, have not yet appeared, in part due to privacy concerns. And of course, library collections have transformed into network services and have become deeply integrated into campus information services. Part of this evolution can be seen in the shifting relationships between libraries and campus information technology planning and implementation departments. In the first age of automation, the library could (and frequently did) stand alone; in the second age, the library became reliant on campus networking strategies; and in the third age, the library is critically dependent on both local-area and wide-area networks and on broad patron access to networked workstations and to network services as diverse as printing and authentication. In the networked information revolution, libraries not only offer their own network-based services but also are becoming increasingly involved in the management and organization of external activities on the network.

This massive range of changes created enormous questions for libraries—perhaps most fundamentally, questions about what constitutes the core of scholarly discourse that they must manage, provide access to, organize, acquire, and preserve and about what constitutes the raw material of future scholarship that must also be collected, organized, and archived. Clearly, this goes far beyond the output of the traditional scholarly publishers and also goes far beyond the concepts of fixed, published, printed works. So much of the new content is outside of the library and outside of the entire system of publishing that it is unclear how much responsibility the libraries can or should take for this material or how they should go about taking that responsibility.[5]

At the same time, academic libraries face difficult problems about how to allocate scarce, increasingly inadequate resources between the present and the future. The traditional published scholarly literature remains of critical importance, and its costs and volume continue to increase out of control. The late-nineteenth- and early-twentieth-century published literature, printed largely on acid paper, continues to disintegrate: there is an enormous need for investments in preservation.

A new set of issues have been added to the agenda: how to describe multimedia digital information effectively and affordably; how to archive digital information; how to address questions of authenticity, integrity, and provenance; and how to structure services around information that needs to be computationally manipulated rather than merely viewed.

Some of the key issues now involve legislation and public policy. Intellectual property rights questions are now clearly one of the greatest constraints to the promise of the digital environment. The issue isn't just paying for currently available digital information products or working out suitable terms for acquiring these products. The difficulties and expense of clearing rights for content, and the complexities of even sorting through the various potential claims to rights in digital content, threaten to block the exploitation of many materials in the networked information world—even though there has been no marketplace for these materials in print for years. And the situation has been greatly aggravated by the recent decades-long extension in the term of copyright (the 1998 Sonny Bono Copyright Extension Act), essentially halting the passage of new materials into the public domain for many years to come.

The networked information revolution has arrived but is still in its infancy. I believe that we will spend the next decade or two refining the technology and building up an ever-growing mass of content. This will include not only new materials that will be created and current resources that are rapidly moving into digital form but also the massive retrospective digitization of our print heritage of past centuries and of our special collections in libraries and in the holdings of archives and museums. The scale and scope of this effort is vast, but so are the benefits to be gained.

Conclusions

Intellectually, automation is easy and rather comfortable, even though it demands considerable management skill and technological judgment to implement. It does not usually challenge fundamental assumptions about roles and missions. The great thing about automation is that you already know where you are heading and what you are trying to do. A very wise person in the 1970s could probably have mapped out the future along the automation timeline through the end of the millennium. Indeed, libraries have always been aggressive adopters of automation technologies—sometimes too aggressive (book-storage robots and ultrafiche are two notable examples). They have been more skeptical and reluctant to adopt innovation (network access, new media, new digital genres, personalization, and reccomender systems are good examples here)—though often they have had good cause for their caution.

Libraries must now turn their attention to defining their missions

A Few Predictions for the Future

I predict that the following issues will be central for academic libraries in the early twenty-first century:

- Establishing a new definition of the "canon" of scholarly communication and the library's relation to it in terms of acquisition or selection, organization and management, access and preservation.
- Addressing the problem of acquiring, managing, and preserving the raw materials for future scholarship as these materials become digital and as they diversify in character.
- Finding a new balance between collective, centralized action and local effort. In a world of shared resources on the network, it is possible to centralize more of the management, organization and description, and preservation of content, and economic considerations encourage such centralization. Yet there are also legitimate needs for local control and for responsiveness to local institutional needs.
- Defining the service boundaries of the library in a world where information is dynamic and is manipulated rather than simply presented to library users. Think about an information resource such as federal census data: in general, this is not simply *viewed* but rather is an object of computational analysis.
- Resolving the systemic funding problems in an environment where costs for traditional materials are increasingly unsustainable and where libraries are simultaneously being confronted with the need to invest in the support of a range of nontraditional networked information resources.
- Developing new roles for the library within the academic enterprise to meet the needs of the networked information revolution. These roles may well include teaching information literacy and information resource evaluation; multimedia authoring and management; partnerships with information-intensive research projects and programs (though it is unclear whether this will involve the library as an institution or a new set of information specialists who are part of the research project teams); support of distance education planning and delivery through the development of new information resources; stewardship of instructional materials, particularly "learning objects"; rights and intellectual property management; and training and consultation in data structuring, representation, organization, and preservation.

and activities in relationship to their transforming context—the information technology revolution in teaching, learning, and research. This will be much harder and more challenging than automation. And it will be a more reactive process: changes in scholarly communication practice, applications of instructional technology, and developments in intellectual property law will shape much of the future of the academic library.

It is striking to me that unlike the progress of automation (modernization) during the past thirty years, which focused on the implementation and management of technology, the agenda for the start of the next century is almost entirely dominated by addressing the effects and implications of technological change. It is, truly, transformation: a basic alteration in the activities of the academic library as an organization as a result of the new technological capabilities and the shifting context of higher education and scholarship.

Notes

1. This history has not yet really been written; for one window into the thinking of the time, see J. C. R. Licklider's book *Libraries of the Future* (Cambridge: MIT Press, 1965). Licklider went on to ARPA and was instrumental in creating the ARPAnet, the predecessor of today's Internet. For a survey that covers some of this early history, see also Clifford A. Lynch and Cecilia M. Preston, "Internet Access to Information Resources," *Annual Review of Information Science and Technology (ARIST)*, vol. 25 (New York: Elsevier, 1990), 263–312.

2. See Clifford A. Lynch, "Beyond the Ordinary Card Catalog: MELVYL Learns from Years of Experience," *EDUCOM Review* 27 (November/December 1992): 20–23.

3.This is most true of historical material. By way of contrast, museums that acquire works by modern artists often get only the work, not the rights to reproduce the work. Using images of the work on postcards or in museum catalogs requires additional negotiation with, and additional payment to, the artist.

4. C. L. Borgman, "What Are Digital Libraries? Competing Visions," *Information Processing and Management* 35, no. 3 (1999): 227–43.

5. Clifford A. Lynch, "On the Threshold of Discontinuity: The New Scholarly Genres and the Role of the Research Library," in Hugh Thompson, ed., *Racing toward Tomorrow: Proceedings of the Ninth National Conference of the Association of College and Research Libraries, April 8–11, 1999* (Chicago: ACRL, 1999), 410–18, online at <http://www.ala.org/ acrl>.