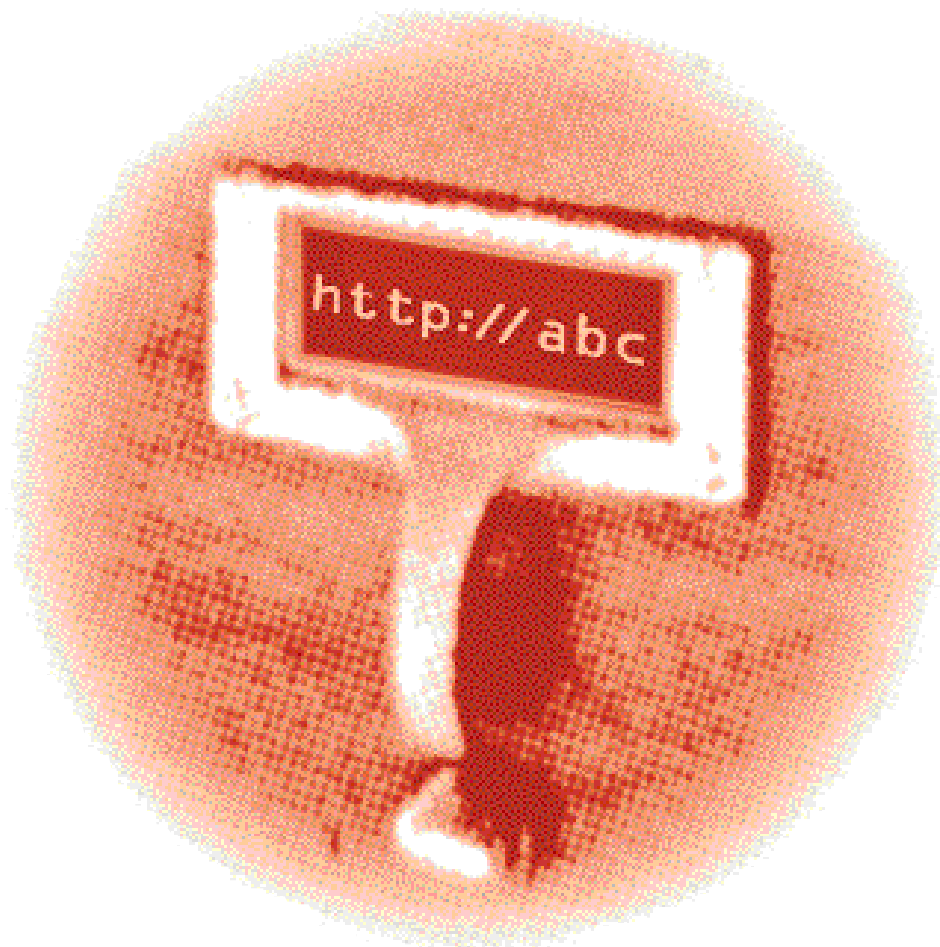

Digital Libraries, **Value, and** **Productivity**

Gio Wiederhold

A digital library is popularly viewed as an electronic version of a public library. But replacing paper by electronic storage leads to three major differences: storage in digital form, direct communication to obtain material, and copying from a master version. These differences in turn lead to a plethora of further differences, so that eventually the digital library no longer mimics the traditional library. Furthermore, a library is only an element in the process of creating, storing, culling, accessing, selecting, and distributing information to customers. While the technical focus of digital library research is on the central functions



Publishing Paths

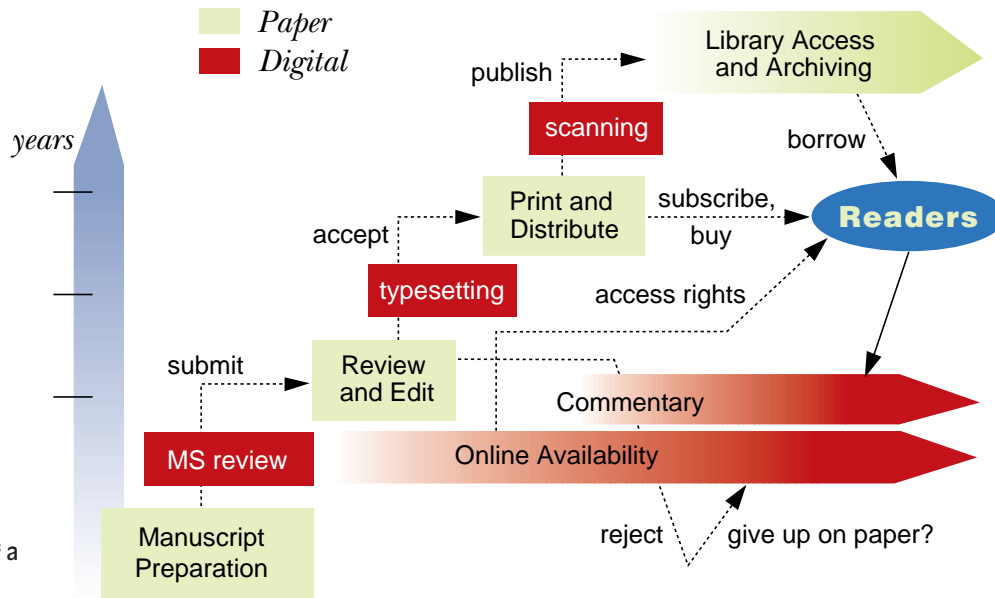


Figure 1.
The path of a
publication

of storage and access, major changes will occur in the interaction within the new systems.

How soon these changes will occur is hard to predict, but the technological foundations for digital library structures are in place and the social and organizational changes will follow. Ignoring the new organizational interfaces being brought about by digital libraries because paper is still dominant is unwise for authors, editors, publishers, booksellers, and readers. After all, printing is barely 500 years old, and made the armies of scribes that existed in the Middle Ages obsolete. The diminished role of monks as keepers of knowledge weakened the dominance of the church and revolutionized the political power structure [14].

We focus on the needs of the customer since services that empower the customer will thrive and systems that frustrate the customer will shrivel. We define customers as individuals who use information to make high-level decisions that affect how products are made, how goods are marketed and sold, how public and private investments are allocated, down to individuals making choices about vacation destinations and movies to watch. Note that these customers are not necessarily identical with the readers of library material. Decision-makers often get their information from staff or consultants who read, cull, integrate, simplify and present alternatives. We have to include these intermediaries in the system model with the librarians and editors; they are equally candidates for displacement as electronic networks and services promote disintermediation.

We are aware of other users who have vital interests in the publishing enterprise, but are not the customers we identified. For instance, academic authors have a

need to publish because it affects their status in the community. Corporate authors may publish to advertise their company's prowess. A substantial volume of writing intends to sway the readers into one viewpoint or the other, and this article itself can be seen as an example. But all these works will only have an effect if they eventually reach a customer: the dean, the purchaser, the allocator of resources. It is quite unlikely for the dean, the purchaser, the allocator to actually read the papers (the dean may not admit that)—they often rely on intermediaries. It is rare that a customer will be affected by one single document, and yet, the aggregate power of the written word is strong. The information path is depicted in Figure 1.

Characteristics of Digital Libraries

A summarization of the critical differences that digital technology brings to information systems follows, so that we can deduce how the ancillary services will be affected.

The initial innovation is that the works (books, magazines, newspapers, records, videos, and the like) are stored in digital form. New formats and standards are being devised to deal with the variety and complexity of the information. To what extent traditional Good Old Paper (GOP) formats should be imitated is still unclear; there is much good in print media and documents, but electronic access will rapidly render many old formats and conventions obsolete. For instance, some electronic mail programs try to imitate the format of paper letters, with salutations and white space for aesthetic reasons. Those features annoy customers who have limited display area on their computer windows.

The second innovation is in communication. Readers no longer need to come to a library; they use electronic methods to have material delivered to them. A work no longer needs to be kept in one place, it will often have linkages to remote subparts and perhaps to executable demonstrations. When relying on digital libraries the important social aspect of congregating in a real library building, where one can learn from the librarian, friends and colleagues, drop children off for a story-telling afternoon, or wait out a traffic jam in a pleasant atmosphere, will be lost. Electronic and virtual surrogates for these social functions are likely to appear, but will no longer be coupled to the library. Giving everyone fair access to digital libraries must be a concern, and linkages must be rapid and economical [20]. Travel on electronic roads can also be clogged. Complex documents, specifically images, occupy wide information lanes, so that traffic and the distribution of digital resources along the highways remain issues. Technologists are working hard to reduce these barriers. The ability to connect to authors and their works directly encourages disintermediation. The roles of publishers, editors, librarians, and research assistants must become bound to online services to survive.

The third difference is that the material is always copied from the master version of the work in the electronic library. That means that the mechanical aspect of publishing and printing an inventory of books or magazines can disappear. The copy need not be a literal image, but can be processed into a composite of items selected from multiple source works. The loss of the mass-production function can cause publishers to disappear. Some authors will not bemoan that loss, and are happy to place their works directly on the networks, so that others can access them without the months and years of delay that are common now.

All these changes will have wide-ranging effects. With electronic publishing and distribution, as soon as the author's fingers have left the keys, the material can be accessed on the World-Wide Web [3]. It is now accessible to anyone linked to electronic networks, which reach an ever-growing portion of the world's population. Readers gain by having rapid access, and can skip the intermediate services of librarians, editors, indexers, and publishers. The citation to [15] is an example.

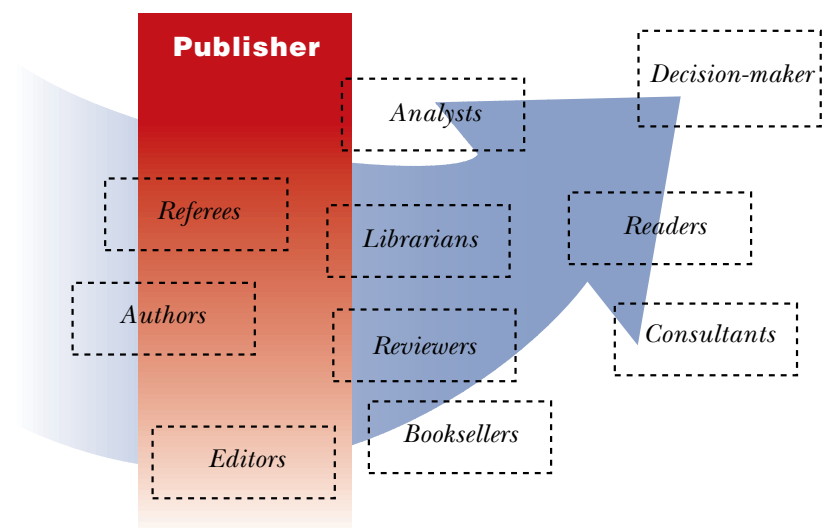
A major issue associated with these changes is the management of copyrights in the electronic world, where traditional concepts, like fixing a work for long-term archival deposit and dissemination, are becoming less relevant [11]. We will focus here on intermediary services.

Lost Services

With the disappearance of traditional publishers, an author will no longer have publishers available as a resource to turn to for advice and income. Without editors, vocabulary use can become more diverse, more original, but also less consistent, making access more difficult. Computer science publishing, being a new field, has always suffered from poor vocabulary control, but we expect direct publishing to invade all disciplines. If editors, librarians, and publishers are bypassed, the customer is deprived of guidance to trustworthy material. The needs for editing, selection, abstracting, and marketing literary works continue to exist, but these functions need new locales and new interfaces.

The services of a volume-editor in a topic, who ranks, assembles, and introduces closely related articles into a collection or a journal has been traditionally bound to mass-publishing, since it is based on the expectation that there will be a sufficiently large number of readers and purchasers of the derived work. Assembling a derived work from material already online is easy and some automatic search robots, such as HARVEST and ALIWEB, build listings of online citations to candidate documents

Figure 2. Participants in the creation, dissemination, and use of information



relevant to a topic [4]. They do not embody domain expertise. It is not a coincidence that a major initiative in electronic publishing is the *Journal of Universal Computer Science* [5] since its founder, Hermann Maurer, sees no need to circumscribe its scope more narrowly.

It is clear that the introduction of digital libraries will severely affect the traditional functions of authorship, editing, publication, critical review, book-selling, libraries, readers, and customers, as sketched in Figure 2. The revolution will not be instantaneous; it is likely to

start slowly, but accelerate over time, keeping many participants unaware until they are trapped by obsolescence. There are many examples of patterns of accelerating change; we recall how trains were affected by airlines, large American cars were affected by imports, and businesses along local roads were affected by the interstate highway network. In each case proprietors of the traditional services remained confident that these effects would be transient or limited, because proprietors presumed their customers really didn't want change. We hear similar arguments today from proprietors of GOP: they point out that managers don't want to lose their assistants, that authors don't want to deal with typesetting, and that refereed archival works represent the essence of scientific progress.

As the redistribution of functions in the digital libraries occurs, all of the participants must consider what their information product is, apart from its physical manifestation, that is, what added information value is generated by their efforts. An author, while viewed as primarily a creative being, also depends greatly on abstracting and reformulating earlier work, especially in scientific areas. An editor makes the work of a good author accessible, and filters out poor work. A publisher advances money to an author, enabling more rapid generation of material, employs reviewers to help with selecting authors' submissions, hires editors, and invests in typesetting, printing, and inventory. Critics assess new publications, and disseminate their judgments to a wide audience. The booksellers focus on market understanding, distribution, and on providing a pleasant environment that encourages people to browse.

The Future in Terms of History

Transmitting information orally and by writing extends back to pre-history, since the concept of history itself is bound to our ability to record events so that later generations can connect their lives to those of their ancestors. We cite three historical analogies.

- The ability to read and write gave priests and kings power. Today, the ability to create and understand computer codes empowers computer hackers. If they are employees, the autonomy they possess often frustrates their managers. Many of them freelance and will use the digital libraries as their means of dissemination. Using electronic distribution, digital gurus can collect followers anywhere, relaxing the traditional binding of influence to land and locality.
- Sharing a spoken language defined communities and sharing a written language defined nations. Today, some computer communities are defined by the computer systems and languages they use.

Fierce intellectual battles are being fought to establish that one language is superior to another, even if they both can express the same computer program, albeit in more or fewer sentences.

Translation will remain a frequent task, not only among natural languages. Contextual assumptions built into languages must be recognized.

- The invention of the printing press broadened access to writing and led to religious reformations and political revolutions. Dissemination of information by facsimile machines has become a method for publicizing political protests. Networks and digital information handling encourage active participation. The technical barriers to becoming a popular author are disappearing, although intellectual ones remain.

We can expect to see new centers of influence and new communities due to digital technology. To be productive in the new world individuals now performing information services will have to rethink their actions. Rethinking means trying to consider the likely locales of future services and who will provide them. There will be new methods and architectures for the information flow. What will be their interfaces to the digital libraries?

Growth

There has been a rapid growth of online access to documents during the past year, and most people have encountered the "information highway" in publicity, if not experienced it in person. Today thousands of individuals and institutions have homepages accessible on the Internet. These pages give information about themselves, from the factual to outright advertisements, and will have pointers to their digital photographs, their digital papers, and their friends and institutions. The rate of growth can diminish due to lack of benefits or access performance bottlenecks. Estimates of end-customer information use are well within the total capacity now available on networks. However, several factors greatly increase this estimate and can lead to poor performance. First of all, we now transmit an order of magnitude more material than we actually read, because the source documents are rarely structured for efficient use, and are often highly redundant. Robotic search programs can place high loads on source sites; new rules of etiquette should be observed [15]. The load increases further when we transmit images of text instead of encoded text and process graphics for information retrieval. But any such bottlenecks will only slow down the changes that digital libraries are initiating.

Services

Table 1 lists services involved in the distribution of information from author to customer, so that we can compare digital methods with those currently in use.

Not all these services are performed for all works. For instance, the computer literature is not indexed using controlled vocabularies, and many works are only edited by their authors. Domain vocabularies or ontologies needed for consistent referencing are costly to develop and maintain, requiring many dedicated individuals in every speciality, and ongoing efforts to establish agreement. Financial pressures have caused publishers to scale back on services in general, and it is often the author who prepares camera-ready copy, with all figures, layout, and suggested index terms included. Although the library and librarians focus on functions in the core of this enterprise, the transition to the digital world affects all tasks. We cannot, in the scope of this article, individually analyze the transition of all the services listed in Table 1. Some of them are already being performed or aided by remote services, and those can serve as models for the evolution of future information processing.

Cataloging

As libraries grew due to print technology, a need evolved to locate books other than by scanning the shelves [1], and librarians began making lists cataloging the books. In 1791 the French revolutionary

government confiscated many private and monastic libraries, and to create a union catalog of their possessions they initiated a collection of reference information by ordering the listing of every individual book on a playing card (reserving aces and deuces for lengthy titles). The cards were filed in drawers, and soon became resources themselves. The cards would be copied, so they could be filed in a drawer arranged by the authors' names, in drawers arranged by title, and in subject drawers. Remington Rand, an ancestor of the UNISYS computer company, made many of these drawer cabinets.

In 1901 the U.S. Library of Congress (LoC) started printing its cards, and made copies available to other libraries. Specific local information was typed onto the LoC cards. Over the years complex cataloging rules have emerged, and "Rule Interpretations" are issued periodically for U.S. librarians by the LoC. The card catalogs became increasingly awkward, as well as targets for vandalism, as more people used them. Today few libraries maintain card catalogs, but the librarians' knowledge continues to be embodied in the digital MARC formats that are being distributed and used for local cataloging. Many libraries today contract their cataloging to outside services; the Ohio State system

Table 1. Functions in the publishing enterprise.

Service	Example	Provider	Support
Writing	this article	author	investment (see Table 2)
Locating citations	reference section	library, colleagues	institution
Selecting works	refereeing	publishers, editors	publishers (pub.)
Editing	spelling, word usage	staff, publisher	institution, pub.
Graphics	figures, icons	staff artist, publisher	institution, pub.
Ontologies	ACM Classification	professional society (PS)	pub.'s editor
Layout	figure placing, white space	publisher's staff	pub.
Composition	creating this topic issue	editor-in-chief	institution, pub.
Printing, binding	a book	printer under contract	pub., vanity author
Master cataloging	assigning ISBN code	Library of Congress (LoC)	government
Storage	source for orders	publisher	publisher
Distribution	physical dissemination	publisher, mailing house	pub., vendor
Local storage	store shelves	bookstore	bookstore sales
Advertising	brochures, journal ads	publisher, bookstore	book sales
Reviewing	critic's columns	publisher, PSs	publisher, PSs
Acquisition	library purchasing	library staff	institution
Local cataloging	university library	library staff, services	institution, LoC
Indexing	preparing MEDLINE entry	funded service	PS, government
Retrieving	getting a book	librarian, bookstore clerk	institution, bookstore
Copying	class use	copy center, staff	institution
Revenue collection	copyright fee	vendor, copyright center,	fraction of sales
Validation	contact users cited	consumer staff	user's institution
Abstracting	executive summary	consumer staff	user's institution
Integration	summary table	consumer staff, consultant	user's institution
Presentation	viewgraphs	consumer staff	user's institution

has spawned commercial services that are widely used. Online access over the Internet is now available, permitting local librarians to rapidly catalog new acquisitions and make them available to readers.

This early use of digital technology illustrates how communication has shifted local work and authority to a service provider. However, the available services focus on books, and much new information worthy of cataloging will not be in book form.

Editing

Manuscript editing, especially of technical material, requires knowledge that may not be locally available. Since many editors are freelancers, some are now providing these services remotely. Falling by the wayside are the traditional proofreader's marks, since their wiggles are hard to encode digitally. However, no new markup standards for editing have yet emerged, so that automated tools to record and enter incremental changes are hard to establish. In direct communication between a single author and an editor standards may not matter, but it is important for a publisher to be able to assess the work contributed by each participant in the process. If there is no publisher to assess and market their work, then editors may wish to have a pointer to their homepage in the marked-up manuscript, to serve as quality reference for subsequent customer inquiries.

Refereeing

A referee invests considerable time in assessing a work to be published, considers its worth in terms of quality and originality, attempts to validate its conclusions, and may request clarification of assumptions that lead to the conclusion. It is rare that an academ-

ically rewarded, if at all. A journal or proceedings may list its referees annually, and a magazine will list its editors on the masthead. Seeing one's name on the homepage of an electronic publication may be less satisfying. The referee's name is rarely accessible to the readers of an individual work. The isolation is intended to reduce bias and revenge, but also reduces the recognition that a good referee deserves. If the work is rejected, no recognition at all is gained. Having early access to new work used to be one of the benefits of refereeing, but with instant electronic publishing most material is already electronically available before it has been refereed. Referees' comments and signatures can be appended to the digital document, directly or indirectly via links to the referees' files, as was foreseen in [21]. Recognition of referees' contributions can provide motivation, and being part of a team generating an important and well-written document can actually be exciting.

Much of the communication for the refereeing process travels via email today, and this speeds up the refereeing process. But further acceleration puts pressure on the referee. Proceedings for conferences have to be refereed within a few months, and the evaluations received by the program committees are often inconsistent, so that the quality of a paper may be voted on, which is democratic, but not a good measure of scientific correctness. Further speed up will reduce the refereeing quality.

If the pool of responsible and thorough volunteer referees dries up, which seems likely, other means of evaluation will have to be found. Changing the reward system seems to be essential, and both recognition and money may have to be provided. However, publishers of academic journals do not have funds to

Falling by the wayside are the traditional proofreader's marks, since their wiggles are hard to encode digitally. However, no new markup standards for editing have yet emerged, so that automated tools to record and enter incremental changes are hard to establish.

ic paper escapes unscathed from the refereeing process, and even popular literature is assessed and modified by professional magazine editors and their consultants prior to release. The outcome of the refereeing process is rejection or obtaining an imprimatur that the work satisfies the standards of the publication and the institution or society sponsoring the publication. The reader of refereed material can have some confidence that the conclusions are not misleading, and that it is worthwhile investing effort in understanding and applying the work.

The effort of the referee is rarely adequately finan-

ce to improve refereeing, especially as subscriptions to paper publications diminish.

In the electronic future it may be best that a work remains visible from the moment an author releases it [9]. Linkages to referees' comments and names may be encrypted so that they are not available to all. Public-key encryption methods can make the annotations available to publication editors or other interested parties, perhaps for a fee. In this approach a referee may become more like a critic, who signs his opinions and becomes a respected or hated participant in the information process.

The author may also add secondary annotations, accepting or rejecting referees' remarks. A new reader can access the revised version without necessarily seeing the annotations. An editor-in-chief of a refereed publication can decide when a paper has received enough analysis and decide to bless the revised paper or ask the author for a revision. An accepted paper can have a further annotation attached. The acceptance is attached to this particular version. Further revisions may be made, of course, and should be encouraged, although the acceptance will remain with the earlier fixed work. Forward and backward linkages will allow readers to navigate through the versions.

Since anyone should be free to attach comments to a paper, the quality of the comments or the commentators cannot be assured. There may be a need for a meta-service, which identifies trustworthy referees and their areas of expertise. That service can be used to filter comments for readers that wish such a restriction—other filtering services can be envisaged.

Indexing

The creation of indexes to bibliographic material is a service that adds much value; it greatly increases the precision and completeness of subsequent retrievals. A prime example of high-quality and comprehensive indexing is that provided by the U.S. National Library of Medicine (NLM). Using a thesaurus (MESH) that is regularly updated, certified indexers at several specialized sites provide entries from the MESH vocabulary for use in the indexes used for NLM MEDLINE systems. The environment provided by MEDLINE is not only helpful when searching for information, it also provides a motivation for consistent use of terminology in medicine and reduces overlapping research and wasted efforts. It is now being extended to cover medical terms from non-bibliographic applications [19]. Most fields in science do not have such quality services; in computer science many redundant or near-redundant terms are used and searching for the right references depends on luck, and is usually best accomplished by interrogating colleagues.

When dealing with multiple domains no consistency of terminology can be expected or even advocated. Since the knowledge in many domains is changing rapidly, the freedom to identify new concepts and bind them to new terms or to redefine old terms has to be supported. Even the efforts made by the NLM cannot assure consistency over subdomains in medicine.

Using the same terms to describe concepts still does not assure consistency of factual information in databases [25]. For instance, the scope assigned to a

term will differ even in closely related domains. As sketched in Figure 3, the list of Employees typically differs between the Payroll application and Personnel. Personnel may include contractors' workers who are not being paid through the payroll, and the Payroll may include retired staff, of little interest to Personnel. Knowledge of the domain context is important in understanding the meaning of terms. Once the domain is identified, perhaps through matching of term usage [18], it can be used also to

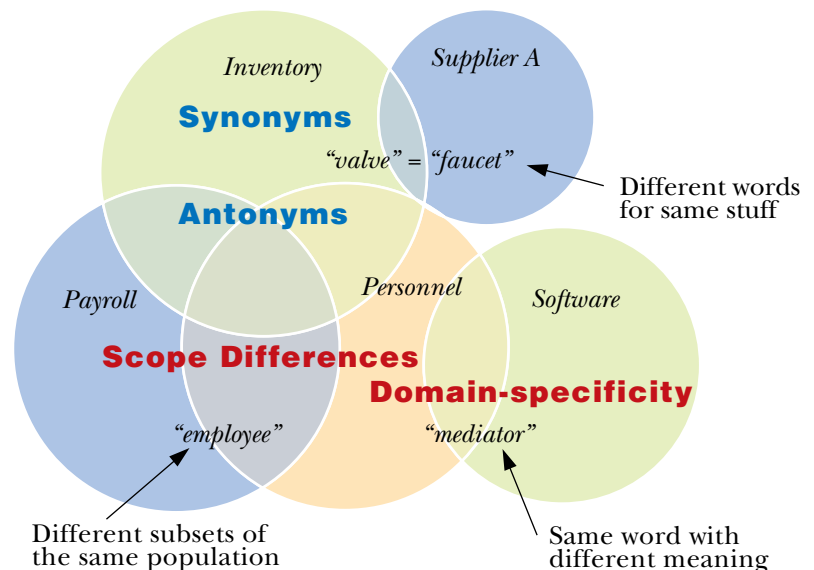


Figure 3. Terminological differences

resolve ambiguous terms, since such terms are typically well defined within a domain.

It does appear that indexing services can be much improved by greater use of standard ontologies and computing-intensive analysis. The domain expertise of the person managing and maintaining such a service adds value to the process.

Result Processing

The processing of retrieved documents, through validation, abstraction, integration, and presentation, is also a candidate for services by specialists. Most end-users need more than rapid access to documents: Decision-makers today depend on their staffs to analyze, abstract, and rank alternatives according to their world view.

To transform data from a library to customer information one would first collect references to candidate documents, using indexes provided by expert services. The quality of documents may be validated, perhaps using referees' annotations as referred to earlier, other citations, such as those collected by the Science Citation Index service; or linkages to demonstrations.

The essence of a document may have to be abstracted for the customer [2]. Automation of abstraction is a difficult problem, and needs domain expertise. Perfection in abstraction is not essential, since digital abstracts will contain embedded references to the source documents. There is success today with specialized documents, using metamodels as a basis. Examples are the creation of comparison tables from K-10 reports submitted by businesses at Price Waterhouse [16], the processing of military messages, and the creation of problem summaries from medical records [7].

Important decisions are always based on multiple sources, and material from many documents has to be integrated. Such integration may best be performed from the abstracts, since they exhibit some commonality of structure. The source documents obtained can be classified and ranked. Eventually a summary abstract needs to be generated, with embedded linkages to all relevant documents and their annotations. Finally, effective presentation may have to be produced. Online presentations can retain those linkages.

Today these tasks are performed by librarians, consultants, and staff. It is unclear how such processing will be partitioned in the future. Speed is likely to be a major concern, and the services will have to compete with the capabilities that the customer can marshal without going to outside services. It is obvious that digital services will all be online, and accessible through networks.

Processing data into information requires that the

service is responsive to the customer. One approach is to design processing methods that are specific to a type of customer, and hope that the market is sufficiently large and that many customers so typed will use the service. A more general approach would be to interpret a model of the customer's needs, and match that against available data to maximize relevance. Such a model of user needs is typically a hierarchy, since feasible solutions to real problems invariably follow a succession of divide-and-conquer paradigms.

New Services

We referred to traditional models for the added-value services that were listed, but in time a plethora of new services can become available.

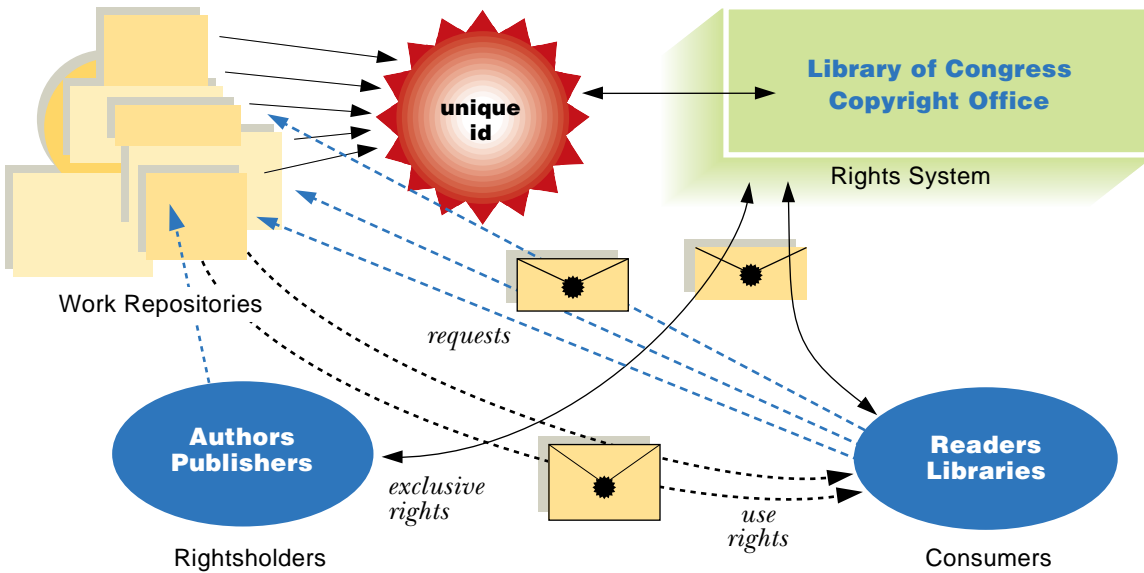
Online Revenue Collection

To support the library enterprise, money has to flow into the system (see Table 2). Where broad public access is desirable, government may provide support, just as it now supports public libraries. Authors typically receive a negotiated fraction of sales from commercial publishers. The publishers receive sales income and copyright fees. For performances of works, fees are collected by organizations as the American Society of Composers, Authors, and Publishers (ASCAP), who distribute the fees to the owners of the rights. While the flow of the revenue stream is torturous, adequate funds do flow through the system to support the publishing enterprise. Many authors and readers are able to ignore the specifics of

Table 2. Current and potential revenue collections paths.

Service	Collection	Distribution	Recipient
Personal investment	author's savings	author	author)
Employer	institutional income	manager's time allocation	researcher, author
Philanthropy	bequests, donations	committee, angel	author
Commercial investment	publisher's advance	publisher	author
Government	taxes	funding agencies	researcher, author
Page charges	publisher	publisher	printer, publisher
Book purchase	bookstore	publisher	printer, author
Book order	wholesaler	publisher	printer, author
Periodical subscription	professional society	academic publisher	printer, publisher
Digital subscription	professional society	publisher	publisher
Periodical copy purchase	publisher	publisher	printer, publisher
Advertising	commercial publisher	publisher	printer, publisher
Periodical article copy	Copyright Clearance Center	publisher	publisher
Article or figure reuse	publisher	publisher	publisher, author
Personal article reuse	free	n.a.	free to the reader
Performance	owner's representative	as ASCAP	author, composer
Digital copying	none	none	free to the reader
Protected digital copying	networked fee collectors	digital publishers	rights holders
Freeware copying	direct by owners	direct by users	rights holders

DL Copyright Management



courtesy of CNRI

Figure 4. Electronic copyright management; adapted from [12]

the system, and have delegated their rights and reimbursements to the publishers.

The services discussed in this article will require fees from customers. The fee for a unit of service must be fairly low, since many services will be needed to support decisions. Several proposals for low-cost, electronic revenue collection have been made; for instance, by the Coalition for Electronic Markets [6], and several firms will now carry online accounts against which charges can be debited.

Protecting copyrights requires being able to determine identity and ownership of a digital work in a reliable way. Having unique identifiers independent of media and format transformations is a problem now being addressed [12]. However, the technological barriers will be overcome if there is willingness by authors, publishers, consumers, and the Copyright Clearance Center of the U.S. Library of Congress (LoC) to cooperate, as shown in Figure 4. It should, in the future, not be necessary to physically deposit two copies of one's work at the LoC to have proof of originality.

The concept of performance of a work is seen by some to match the action of making a digital copy for oneself, leading to adaptation of the existing concepts for revenue collection from performances. The current copyright law does not require payment for performances in certain deserving venues, such as veterans' hospitals—applying these rules to a digital world will not be easy.

Proposed changes to the copyright laws to cover digital copying broaden the restrictions to enable protection using current technology [23]. But these

restrictions go beyond current practice and engender opposition [13]. At the same time the European Mathematics Society presents arguments in favor of free access to electronic publications, including author's preprints of all works submitted to be published in paper journals, leaving to publishers and libraries the costs and income of making paper copies. Their report states that the costs of services other than writing, refereeing, and printing "are small" [9]. In that view, a small community can acquire its limited number of readers without the interposition of a publisher.

The ACM plans to provide access to its material free for its paying subscribers. The fixed, final copies will be held by ACM but be made available for downloading and linking. A number of details must still be resolved. For instance, the ACM, and many other professional societies have multiple classes of subscribers, including members of only a single Special Interest Group. The authentication of membership for all these categories is a technical hurdle not yet overcome. Non-members may have to pay copyright fees. The success of this approach will depend on the quality of the access service, so that authors and their institutions will find it convenient and effective to delegate digital distribution.

It is crucial that digital library services such as cataloging, editing, refereeing, indexing, and processing are designed to enable online revenue collection. Most current prototypes of digital libraries are supported by

the government, and many of the source documents are public and their use is free. Refereeing is provided using the traditional, unreimbursed approach. It will be important that future experiments include revenue collection, although customers may be unhappy to pay for what used to be free. Free services are expected when customers deal with government data, but it is unlikely that quality value-added services using such data can be delivered without cost to the recipient.

Image Search

Pictures, graphics, and photographs are important representations of information. Images collected in libraries are now located by specifying some external parameters; for instance, satellite images can be located by naming the spacecraft, the observing instrument, the orbit number, and the longitude. This meta-information says little about the image content itself. Locating images instead by their content (i.e., forest areas that are being clear-cut, or x-rays showing certain abnormalities) requires much knowledge, and the supporting programs are likely to be candidates for improvement for many years to come. Image-based retrieval services can deliver substantial added value for users that access image data.

To search for images by their content the image itself has to be processed so that the contained objects can be identified. A two-stage approach seems viable for automation, and early products are appearing [10]:

- The images in the library are analyzed to recognize objects or features of objects. Features used include color distributions, texture, line-segment patterns and orientation, basic shapes as circles, rectangles, and the like. The features are captured in comprehensive indices for each image.
- When the image file is to be searched, the requester shows or sketches an example of the object. A corresponding search index key is created, and may be edited. Retrieval using this key generates a ranked list of images according to the match of key features versus file indices [10]. For instance, houses seen from space are mainly composed of rectangles that have areas of about 10 to 1,000 square meters and edge ratios from 0.2 to 5.

The definition and selection of features, and the weight to be assigned to them, require the insight of domain specialists. For instance, a search for agricultural features in multispectral images obtained from space will differ from a search for clothes in a sales catalog. Here is again the potential for providing an online service, since image libraries will comprise an important fraction of future digital libraries.

Dynamic Books

Many books are prepared by multiple authors, and composed for presentation. Textbooks, even those

written by a single author, consist of distinct chapters, and are organized so that a student or teacher can skip those chapters that are not relevant. A paper book that contains all possibly relevant material gets heavy, costly, and awkward. Some publishers (McGraw-Hill, for example) now allow such books to be composed from selected chapters to order, specifically those destined for classroom use.

It is a small step to move to dynamic books, where a skeleton designating the candidate chapters is pre-defined, but the actual chapters are obtained from digital resources when needed. The chapters would be latest versions, and convey the most-up-to-date knowledge. Some anticipatory planning will be required to assure consistency. A cross-reference to a chapter that is not included has to become a remote link, to be followed only if the reader wants to explore outside topics. The table of contents and the index of a dynamic book might include all candidate chapters, with indications to distinguish chapters seen and optional chapters. The bibliography may only collect entries of chapters that have actually been obtained.

As the source chapters are revised, dynamic books can remain up-to-date to an extent that is not feasible with paper books. For revised chapters, some cross-references may be confusing. Including timestamps with such references can help, and enable the reader to go back to past, archived versions of the chapter.

Unless the reader controls the entire composition of a dynamic book, there are now two levels of authors, the editing author, who defined the range of candidate compositions and is responsible for coherence and completeness of the material and the linkages and the base authors, who have written, and maintain, the individual chapters. The editing author and the base author both contribute, and deserve intellectual and financial rewards for their efforts. The traditional copyright mechanisms need to be adapted to digital libraries and the new dissemination methods that are enabled by the digital highways. Revenue collection has to deal with the diversity of contributors.

Rural Access

There will still be the underserved, for whom the free libraries provided relief in the past. People who live far from the main "information highways" now pay high access charges, in the form of rural telephone tolls, which can cost more to reach the nearest town than the price of an interstate connection. In the 1930s the Rural Electrification Service Administration helped bring light and appliances to every farm. Its bureaucracy still exists in the Department of Agriculture and spends over \$140 million per year. It might find a jus-

tification for its existence in repeating its success in the next century by disseminating information rather than power to rural communities.

Conclusion

The functions of a library are to acquire works, store them, make the works available to the reader, and reimburse the author and the publisher for their efforts. Expected changes of these functions as digital library technology takes hold have been discussed, but no clear picture exists of the future nor of its arrival time. Changes in technological capability are rapid, and are being adopted in a piecemeal manner. Government support for digital libraries can only help in providing some examples of new technology, to be adopted if their use will improve quality or broaden access, but cannot guarantee acceptance or conformance.

Today computer storage, processing capacity, transmission, and presentation of literary text greatly exceeds the capabilities of previous methods. The entry of pre-existing documents remains tedious. Most current works are prepared using computers, so that recent material can easily be made available. Many of these works are being created with linkages, and past, unlinked works will become rapidly unused. Important past works will be entered and stored somewhere on the net and become available online.

The publishing enterprise will experience the greatest disruption. Its members will only be able to survive if they learn to focus on the value they are providing. Those values are likely to be domain-specific as well [17]. Today many publishers are responding to calls for more scientific information by increasing the number of publications, and raising the prices of the journals because of fewer subscribers for each of these publications. The effects of tight budgets and higher costs force many libraries and individuals to cancel subscriptions and have initiated a vicious spiral where in the end only a few paper copies of a work will be sold, and the remaining distribution will be by copying and interlibrary loan. Interlibrary loan is also simplified through the use of networks [8]. Given this situation, independent digital library services may flourish.

Digital revenue collection will become routinely available, although its formats and distribution paths are still quite unclear. The alternative, having free services, will invariably lead to overuse, imbalances, and access restrictions based on allocation rules. The experience of

many disciplines, namely that dealing with overuse leads to reduction of quality, is likely to be true here as well.

Commercial publications, supported mainly by advertising revenue, will feel the pinch from other influences. Some advertising revenue is already flowing today to the few cable shopping channels. When these services become interactive, allowing customers to browse through domains of interests, ask specific questions on, for example, quality assessment, from an independent source, see demonstrations online and the like, then revenues of GOP service publications will be severely affected. The only advertisers that may remain are perfume makers, with their printed odor-strips, for which I have not yet seen a digital replacement.

Problems to be faced will include compliance with standards and protection of intellectual property as works are composed of widely distributed linkages, and value-added services are interposed at many intermediary stages. Publishers now perform a function of process-management and cohesion, tied to a paper-based publishing process. If they can adjust, the same role may be viable in the electronic world, where no paper master copies exist.

Improved response times will be expected throughout the system. Calling in consultants as intermediaries for selection, editing, refereeing, and similar tasks and obtaining their output weeks or months later will not be feasible. The issues and possible solutions presented here have been discussed for some time, but the rapidly increasing use of networks for information access may make new services economically viable. New opportunities will develop for innovators who can provide responsive and high-quality online services, and who learn to use electronic mediation tools effectively [24].

It remains hard to predict how rapidly the change to intensive use of digital libraries and other network services will occur. With responsive and economical services providing quality information, making customers more productive, the growth of online information use will continue.

Acknowledgments.

I have to thank many colleagues for contributing concepts and background to this article. The development of the Digital Library program at NSF, with the cooperation of colleagues at NASA, ARPA, and CNRI provided many insights, motivations, and opinions. Peter Denning, Oscar Firschein, Ed Fox, Rebecca Lasher,

The effects of tight budgets and higher costs *force many libraries and individuals to cancel subscriptions and have initiated a vicious spiral where in the end only a few paper copies of a work will be sold, and the remaining distribution will be by copying and interlibrary loan.*

Giuliana Lavendel, and an unknown referee provided important editorial feedback. The unknown referee had insightful comments and deserves the recognition which this article cannot appropriately give. Opinions presented here are strictly my responsibility, and I reserve the privilege to change them as I learn more about libraries, librarians, and publishers. **□**

References

In this electronic world it is to be expected that many references will be digital as well, some citations here are only available online.

1. Baker, N. DisCards; *The New Yorker Magazine* (Apr. 4, 1994), 64–86.
2. Berleant, D., and Berghel, H. Customizing information. *IEEE Comput.* 279, 10 (Sept./Oct. 1994) 96–98; 76–78.
3. Berners-Lee, T., Cailliau, R., Groff, J., and Pollermann, B. World-Wide Web: The information universe. *Electronic Networking: Research, Applications and Policy* 1, 2 (1992) 52–58.
4. Bowman, C. M., Danzig, P. B., Hardy, D. R., Manber, U., and Schwartz, M. F. The HARVEST information discovery and access system. In *Proceedings of the Second International World-Wide Web Conference* (Chicago, Illinois, Oct. 1994), pp. 763–771. <ftp://ftp.cs.colorado.edu/pub/cs/techreports/schwartz/Harvest.Conf.ps.Z>
5. Calude, C., Maurer, H., and Salomaa, A. J.UCS: The Journal for Universal Computer Science and its Applications to Teaching; Symposium Didaktik der Mathematik, Klagenfurt, Schriftenreihe Didaktik der Mathematik, 1994.
6. Cox, B. Superdistribution. *Wired*. Idées Fortes section (Sept. 1994) <http://www.site.gmu.edu/bcox/CEM>
7. deZegher-Geets, I., Freeman, A. G., Walker, M. G., Blum, R. L., and Wiederhold, G. Summarization and Display of Online Medical Records. *M.D. Comput.* 5, 3 (Mar. 1988) 38–46.
8. Dutcher, G. A. DOCLINE: A National Automated Interlibrary Loan Request Routing and Referral System. *Inf. Technol. Libraries*, 8 (1989) 359–370.
9. Charta of Free Electronic Access to Publications. European Mathematical Society. (Dec. 1994) <ems-e-jour-committee@esi.ac.at>.
10. Faloutsos, C., Barber, R., Flickner, M., Hafner, J., Niblack, W., Petrovic, D., and Equitz, W. Efficient and effective querying by image content. *J. Intell. Info. Syst.* 3, Kluwer Academic Publishers, 1994, 231–262.
11. Garrett, J. R., et al. Toward an electronic copyright management system. *J. Amer. Soc. for Info. Sci.* 44, 8 (Sept. 1993), 468–473.
12. Kahn, R., et al. Proposal for Electronic Copyrights Management. Corporation for National Research Initiatives, Reston, VA, 1993.
13. Kaman, R. Act Now to Preserve Information Rights. O'Reilly and Associates, Sebastopol CA. <http://gnn.com/meta/imedia/features/copyright>.
14. Kilgour, F. *The History of the Book*. To be published.
15. Koster, M. Rules for Web Robots. Nexor, 1994. <http://web.nexor.co.uk/users/mak/doc/robots/robots.html>
16. Krishnan, R., and Steier, D. H. Adaptive access to a digital library of corporate information. Digital Libraries Workshop. Rutgers Univ., Newark, NJ May 1994. <http://www.cs.umbc.edu/conferences/dl/abstracts/krishnan.html>
17. Lederberg, J. Keynote presentation at CAMIS. April 29, 1993, Stanford University.
18. Li, W.-S., and Clifton, C. Semantic integration in heterogeneous databases using neural networks. In *Proceedings of the Twentieth Conference on Very Large Data Bases (VLDB)*, Morgan-Kaufmann, 1994, pp. 1–12.
19. Lindberg, D. A. B., Humphreys, B. L., and McCray, A. T. The unified medical language system. *IMIA Yearbook of Medical Information* 93, Schattauer Verlag, 1993.
20. Maurer, H. Advancing the ideas of World-Wide Web: Hyper-G. In *Proceedings of the Distributed Multimedia Systems and Applications*. Honolulu, 1994, pp. 201–203. <http://iicm.tu-graz.ac.at/pub/Hyper-G>.
21. Nelson, T. *Computer Lib—Dream Machines*. Distributors, South Bend, IN, 1974.
22. Faloutsos, C., Barber, R., Flickner, M., Hafner, J., Niblack, W., and Petkovic, D. Efficient and effective querying by image content. Special Issue on Integrating Artificial Intelligence and Database Technology. *J. Intell. Info. Syst.* 3 (1994).
23. Samuelson, P. The NII intellectual property report. *Commun. ACM* 37, 12 (Dec. 1994) 21–27.
24. Wiederhold, G. Mediation in the architecture of future information systems. *IEEE Comput.* (Mar. 1992) 38–49.
25. Wiederhold, G. The roles of artificial intelligence in information systems. *J. Intell. Info. Syst.* 11, 1 (1992) 35–56.

Announcing ACM's Online Publications Catalog

ACM, the first society in computing, is proud to announce its Online Publications Catalog, accessible via the World Wide Web (WWW). This graphically pleasing catalog highlights all of ACM's journals, books, proceedings, videos and CD-ROM products.

Join us online any time of the night or day!

How To Get To Us:

The WWW URL is:
<http://info.acm.org/catalog/>

For more information contact the ACM Member Services Department:

Phone: 1-800-342-6626 (US & Canada)
+1-212-626-0500 (Metro NY & Outside US)
+32 2 774 9602 (Europe)



About the Author:

GIO WIEDERHOLD is a professor of computer science at Stanford University. He also holds courtesy appointments in Medicine and Electrical Engineering. He is active in the application and development of knowledge-based techniques to database management, information systems, and software construction. **Author's Present Address:** Department of Computer Science, Stanford University, Stanford, CA 94305-2140; email: gio@cs.stanford.edu. His home page is at <http://www-db.stanford.edu>.

Permission to copy without fee all or part of this material is granted provided that the copies are not made or distributed for direct commercial advantage, the ACM copyright notice and the title of the publication and its date appear, and notice is given that copying is by permission of the Association for Computing Machinery. To copy otherwise, or to republish, requires a fee and/or specific permission.