Electronic Commerce and Digital Libraries

Andrea L. Houston and Hsinchun Chen
ISDS Department, Louisiana State University
Baton Rouge, LA 70803, (225) 388-2507
ahoust2@lsu.edu
Management Information Systems Department, University of Arizona
Tucson, Arizona 85721, (520) 621-4153
hchen@bpa.arizona.edu

In this chapter we discuss digital libraries from an electronic commerce perspective. The focus is on what the two have in common. The first section is an introduction which discusses some of the impacts that digital libraries and electronic commerce have had on our lives. The second section discusses common driving forces behind the two. The next section discusses common challenges, with an emphasis on the digital library perspective. The fourth section discusses several common issues, in particular, social, legal, quality, security and economic issues that both digital libraries and electronic commerce must address. The discussion in the fourth section primarily presents a digital library perspective, although the issues are important to both digital libraries and electronic commerce. Finally, the chapter closes with a conclusion.

(Digital Libraries; Electronic Information; Information Retrieval)

1. Introduction

Advances in information technology and information management are dramatically changing our lives, especially the way we conduct business (e.g., electronic commerce) and locate and access information (e.g., digital libraries). Information, as discussed in this chapter, should be thought of in the broadest sense including at a minimum, text, numeric, image, video, and audio forms.

Exploiting information-related advances has become a strategy that few businesses can afford to ignore. For example, two years ago very little business was conducted over the Internet. Today an informative and engaging web presence is becoming a competitive advantage that organizations are aggressively pursuing. Consumers can locate information about a company or product using the World Wide Web (WWW). They can also order products and services through web-based catalogues and shop in web-based virtual shopping malls, browsing and purchasing products and services from more than one
corporation. A variety of financial services (stocks, loans, banking, bill payment, etc.) are also available over the WWW.

Eventually, even highly personal and interactive services such as medical diagnosis and treatment recommendations for house-bound patients will be available on-line. It is not unlikely that the next generation will live in totally wired homes verbally interacting with intelligent virtual agents who will manage their homes and their lives. These agents will automatically perform such services as: ordering groceries and other products; scheduling services; providing reminders for appointments and errands; paying bills and handling other financial tasks; filtering mail; recommending articles and news items to read or (more likely) view; recommending and scheduling entertainment; performing background checks on child-care providers and other home-oriented service providers (gardeners, pool maintenance, repair and remodeling, cleaning, catering, etc.).

The location and provision of information traditionally performed by libraries and librarians has also dramatically changed. Clients no longer need to leave the home or office to locate and access information as it is usually available on-line via a digital gateway provided by a library (or other information provider, e.g., electronic publishers or individuals) (Wiederhold, 1995). Clients are no longer limited to information physically available in the nearest library. Today, through inter-library agreements, they have access to information from a wide variety of globally distributed information repositories (sources). Eventually, we will be able to take virtual tours of museums, historical sites and natural wonders as well as attend virtual concerts and theater performances, or watch an variety of types of movies, in addition to reading, viewing or listening to books, articles, lectures and music, all through our local library gateway.

What do electronic commerce and digital libraries have in common other than the fact that information technology advances have dramatically changed them? For one thing they both are involved with “the creation of information sources and the movement of [that] information across global networks” (Adam et al., 1996). This means that many of the technological changes and economic pressures driving the explosive growth and changes are the same. Similarly, both areas are concerned with the identification and delivery of relevant information to interested customers. This means that they face many similar issues, in particular similar economic, social, legal, quality and security issues. In this chapter we will explore some of these similarities, including the driving forces and the issues confronting both the providers and the users of electronic information.

2. Common Driving Forces

Electronic commerce and digital libraries involve organizations meeting the information needs of their clients via electronic information and global communication networks. To do so effectively requires a combination of technological advances and the ability of an organization to design, construct, manage and use global electronic networks (Mansell, 1996). Organizations must also be able to rapidly adapt to dynamic changes in technology and to cope with the size, scale and complexity of both the networks themselves and the information available through them (Atkins, 1997). Flexibility, nimbleness and creativity are important abilities to differentiate an organization from its competitors.
Not only is the information exchange medium changing from physical objects to networks, but the information itself is changing which in turn has an impact on the medium. Traditional information exchange was textual or numeric, involving simple, standard data types. Electronic information exchange now includes audio, image and video data as well as a need for "graphical" information (i.e., the continuous output of an EKG machine). These data types require more bandwidth for transmission. Furthermore, there is no single internationally accepted transmission standard for any one of them often making information exchange challenging.

2.1 Common Goals

Both electronic commerce and digital libraries can be thought of in terms of a common generic model proposed by Adam et al. (Adam et al., 1996): “providers (information providers, merchants, retailers, wholesalers) make multimedia objects available to consumers (customers, information seekers, users) in exchange for payment.” A generic system for either electronic commerce or a digital library could be characterized as “a collection of distributed autonomous sites (servers) that work together to give the consumer the appearance of a single cohesive collection” (Adam et al., 1996). In practice, each site will most likely store a large number of information objects in a wide variety of formats on a wide variety of storage media (Lynch and Garcia-Molina, 1995). Consumers accessing the information will have a wide range of expertise in key access-related areas such as computer literacy, collection navigation abilities, and domain knowledge (Atkins, 1997).

Electronic commerce focuses on business-related interactions and transactions that serve the needs of consumers, sellers, intermediaries and producers of goods and services. Digital libraries focus on interactions between information producers (authors, publishing companies, government organizations), “librarians” (information locators, indexers, and filterers), and information seekers. Digital libraries use electronic commerce techniques when determining appropriate pricing and cost models for their services. Electronic commerce uses digital libraries as information repositories (Adam et al., 1996; Kalakota and Whinston, 1996; Lynch and Garca-Molina, 1995).

One of the main goals of any organization involved in the information business is to add value to information that it provides or manages. The most common kind of added value is access. Access value either improves access time (retrieval speed and/or timeliness), availability (recall), content (relevance) or some combination of the three (Atkinson, 1996). Historically, organizations focused on generic improvements. The current trend is to customize or tailor improvements to information access at the individual client level.

2.2 Advantages of Electronic Information

The advantages of electronic information for digital libraries and electronic commerce are:

- **Access** - Electronic information is more readily available to customers (assuming that they have devices to generate, transmit and receive it) as there is no (or a dramatically reduced) purchase and delivery cycle. Customers don't need to travel to access information as it is always available on-line. There also is less effort for
the information provider. Electronic information need only be created and stored once to be immediately available over a network simultaneously to multiple users as opposed to multiple copies being generated over time and provided to customers via traditional (manual) distribution channels (Kalakota and Whinston, 1996; Lesk, 1997; Reddy, 1996).

- **Flexibility** - The content of a single item of electronic information can change frequently using just one copy, something not possible in non-electronic media which requires modifications and updates to be republished as revisions or new editions. While this allows errors to be corrected and the most up-to-date information to be available, it can cause problems for archiving, and differentiation between multiple versions and authors. One way of addressing this issue and increasing flexibility is by creating links to modified and related information via hypermedia and hypertext functionality. Another approach, taken by Wilensky (Wilensky, 1996) (Berkeley), is the multivalent document approach in which “complex documents comprise multiple layers of distinct but intimately related content”.

- **Economics** - Electronic information is cheaper to produce, store, distribute, and reproduce or copy. Furthermore, information providers can cooperate with each other by providing a gateway or links to information managed or provided by others. This allows providers to specialize on certain kinds of information thus conserving acquisition and production budgets while still providing access to a wide range of information (Lesk, 1997).

- **Preservation** - Electronic media does not disintegrate as readily as other types. Paper is a particularly vulnerable medium as it is susceptible to the problems of acid paper and binding disintegration as well as destruction through innocent physical handling and vandalism. Other kinds of media (including tape, images, negatives, vinyl records, etc.) are susceptible to disintegration due to pollution, catastrophic events (floods, and other natural disasters), humidity, light, insect and other kinds of pests, mold and mildew, vandalism and human handling.

Nonetheless, there are advantages to paper and other non-electronic forms of information. Some of these include (Kalakota and Whinston, 1996; Lesk, 1997):

- **Economics** - While electronic information may be cheaper to produce, store, modify and distribute, it is still much easier to determine a fair or market price and cost for a physical object or copy. To-date, there isn't a commonly accepted economic model that can accurately and fairly determine either costs or prices for electronic information.

- **Ownership Issues** - When information providers own a physical copy, decisions about acquisition and archiving are relatively straightforward. If information providers only own a link or gateway connection to the information certain kinds of problems arise (Feldman, 1997). For example, if an information provider decides to cancel its subscription to regularly published information (such as a journal) how will access be controlled? Obviously access to future issues should not be permitted but the right to access past issues has already been negotiated and paid for. Dynamically keeping this kind of information accurate increases the record-keeping and access control policies, procedures and processes for
information providers. What should be done about an information provider that goes out of business or an information item that goes "out of print"? In both cases the information provider can no longer afford (or wish) to support physical storage. How can the rights of owners of links to that information be protected?

- **Copyright Protection** - With physical copies it is easier to track copyright infringements (at least there is physical evidence for prosecution purposes). Copyright violations are more difficult to track when anyone with on-line access can download information, plagiarize or modify it, and upload it making the "new" version available to others.

- **Quality Control** - The publication business (especially reputable publishers, editors and other kinds of information providers) lends credibility to information *content*. Consumers are less concerned about fraud, plagiarism, and unreliable or invalid information with non-electronic formats. Most non-electronic information (especially scholarly information) is subject to some kind of peer review, editing or similar validation process, further augmenting perceptions of quality.

- **Stability** - Physical copies also provide a kind of stability to information, as changes or modifications cannot be made except in revisions or new editions which are easily distinguishable from the original. Similarly, it is easier to distinguish multiple versions of information as well as the contributions of multiple authors and reviewers on physical copies.

- **Volume Control** - Traditionally, customers limited the volume of information they searched physically - only so much information was physically accessible. In cases of excessive volumes of available information, the advice of an expert was often sought. Now that anyone can become an information provider, the volume of information available has dramatically increased and will continue to do so. Even domain experts cannot keep up with this explosive growth in volume, making it more important for information consumers to filter out irrelevant information and information of questionable quality.

- **Quality of Experience** - Not everyone prefers to interact with a computer in order to get information. Many express strong preferences for "curling up with a good book", an emotion few express towards a computer. The quality of a computer sound system can rarely compete with the quality of a high-end home sound system or the acoustics of a finely engineered music hall or theater. The same is true of the image quality available on computer screens. Most people would agree that computers in general are at the low end of the scale for viewing images and movies.

### 2.3 Economic Pressures

Although the increase in electronic information is partially driven by economics, the economic pressures are slightly different for electronic commerce and digital libraries. There is an obvious economic advantage to both of them from a storage and preservation perspective. Electronic information is less expensive to store and easier to maintain. There is also an obvious economic benefit from the ability to provide a previously unattainable level of service, i.e., "individual words and sentence search and delivery of information to the user's desk-information that does not decay with time, whether it is words, sounds or
images” (Lesk, 1997). In addition, previously information that was either unavailable or difficult to attain is now readily available electronically (i.e., large government collections).

In the case of electronic commerce, Malone (MIT) argues that there are three main economic effects enabled by technology (Mansell, 1996):

- **Communication effect** - increased amount of information exchanged at a reduced cost;
- **Brokerage effect** - more effective match between buyer and seller; and
- **Integration effect** - move effective stage coupling in the value-added chain.

While digital libraries are subject to the economic benefits of the communication effect, there are other economic pressures unique to them including:

- **Inflation** - an extremely rapid rise in library operating costs (especially in acquisition or collection expansion, particularly of scholarly journals). In the past 20 years, journal prices have soared by 400 percent while book and monograph prices have increased by 40 percent (Frye, 1997);
- **Volume** - an explosion in the amount, variety and complexity of available information;
- **Maintenance** - the preservation crisis in existing collections, especially with regard to acidic paper (nationally the replacement cost of disintegrating print materials extrapolates to approximately $35 to 45 billion (Frye, 1997));
- **Multimedia** - the increasing amount of information available in multimedia format which requires special viewing or listening facilities. The cataloging and storage requirements for multimedia information are also different from traditional print information;
- **Collaboration** - the advantages (both economic and improved level of service) from resource sharing among both libraries and other information providers;
- **Timeliness** - as electronic information is easy to produce, distribute and duplicate with few of the costly problems of multiple handling and redistribution, it encourages not only a dramatic increase in cost efficiency, but also a dramatic savings in time (Ginsparg, 1997); and
- **Scholarly Communication** - experts agree that the most significant economic factor driving digital libraries is the severe cost problem associated with scholarly communication (Drabenstott, 1993), in particular the excessive cost of providing access to an appropriate number of scholarly journals (Atkinson, 1996; Frye, 1997; Lesk, 1997) to maintain an adequate level of service for customers. For example, according to Andrew Odlyzko (Bell Labs), “a good mathematics library spends $100,000 per year on journal subscriptions, plus twice more on staff and equipment. [The] US spends as much money buying mathematics journals as NSF [the National Science Foundation] spends on mathematical research” (Odlyzko, 1996).

### 2.4 Technological Advances

There are many technological advances in information production, management and distribution that are responsible for enabling the increase in electronic information. They are too numerous to describe in detail but include such things as advances in: 1) storage media; 2) digitization or information capturing techniques (i.e., OCR technology); 3)
automatically indexing and organizing large volumes of information (Schatz and Chen, 1996); 4) computing speed; 5) network technology (including data compression); 6) content-based search and retrieval (Schatz et al., 1996); 7) feature-based or texture-based search and retrieval (Smith, 1996); 8) data compression and full-text indexing (Witten et al., 1994); 9) resource or knowledge discovery; 10) multimedia and hypertext; 11) standards (i.e., Standardized General Mark-up Language (SGML), and Hypertext Mark-up Language (HTML)); 12) visualization techniques (Rao et al., 1995); 13) object-oriented techniques; and 14) improvements in user-interface design (Drabenstott, 1993; Kessler, 1996; Lynch and Garcia-Molina, 1995).

3. Common Challenges

Information providers have always been concerned about the quality (accuracy, completeness and timeliness) and the availability of information (ease of access, ease of use, timeliness and preservation) (Kessler, 1996). From a technical perspective, common challenges for electronic information fall into three general categories: 1) information acquisition and storage; 2) techniques to identify (locate), retrieve and filter relevant information from vast collections; and 3) providing universal access from a wide variety of information sources to a wide variety of information seekers.

3.1 Information Acquisition and Storage

The first major challenge for electronic information is that most information is not digitized. This is particularly true of existing information. In the future, policies can be established to digitally create new information, but the immediate challenge is to determine how to cost-effectively convert non-digitized information. Conversion challenges include both a technological component (e.g., improving OCR technology, advances in storage media) and an evaluation component (e.g., selection and prioritization of information to convert, choosing the appropriate level(s) of digitized quality - critical for images). Other related challenges are identifying methods to capture and index continuous media in real time and techniques for processing, storing and managing vast volumes of extremely complex electronic information (Adam et al., 1996; Christel et al., 1995; Waclaw, 1996).

The ultimate goal is not to keep a pristine, stagnant historical repository, but to encourage information use and reuse by interested consumers. This requires some kind of universal, and probably automatic (due to the volumes involved) indexing or feature identification mechanism for all data types that can be used to classify information for later extraction. At a minimum, electronic information metadata (information about the information features) needs to include feature relationships (for example, spatial, temporal and other domain-dependent logical relationships) and the ability to describe non-numeric and numeric features (Smith, 1996). Much of the research on feature extraction has been done in the area of image databases and video indexing.

Another major problem is the dynamic nature of electronic information (Huser, 1995). The content can change over time requiring either multiple copies or versions to be individually stored, or some mechanism that allows version differentiation for a single copy (Levy and Marshall, 1995). Electronic information facilitates multiple authorship and collaboration (integrated or leveled). Consumers need to be able to easily differentiate the contribution of each author.
Information collection managers need to develop mechanisms to determine what information and which version(s) to include in their collections. This can create information ownership and archiving issues. For example, if an information provider decides to only keep a pointer or link to a certain piece of information, then what happens when the owner(s) or manager(s) of the electronic copy decide it is no longer cost-effective to keep it? What are the responsibilities of the owner to notify pointer or link owners of changes or deletions (a classic problem on the WWW)?

3.2 Relevant Information Identification and Filtering

Assuming that information providers and managers can solve the acquisition and storage challenges, the next set of challenges involve finding ways to make the right information available to the right customers at the right time to improve decision making and maximize the utility of the information (Kalakota and Whinston, 1996; Lynch and Garcia-Molina, 1995). Not only must customers be able to identify or locate information potentially relevant to them, but they must have some way to either filter information so that only the most relevant information is returned and organize (via a ranking or categorization scheme) volumes of potentially relevant information into manageable units. Most users do not have the time to wade through vast amounts of information looking for the critical key informational gems. It is likely that intelligent artificial agents will be heavily involved in coping with the challenges of information location and filtering (Atkins et al., 1996; Knoblock et al., 1996).

There are at least two different kinds of information location processes. The first kind is useful in a broad-based search where the information need has not yet been narrowly defined. Recall is most important in this kind of search. Due to the volume of information returned, effective organizing and classifying techniques must be provided. Relevant information will probably be broadly dispersed among several distributed heterogeneous information sources. Therefore, the key challenge will be to present a seamless information integration or combination to the customer. Customers interested in this kind of search will probably want the information to be summarized for quick perusal (Schatz et al., 1996). Many of the problems and solutions associated with this kind of challenge have been well documented in the heterogeneous and federated database literature.

The second kind of information identification process involves a very narrow, well-defined and focused search. This kind of search requires very detailed information, most likely from a single information source. Precision will be most important and therefore effective filtering techniques will be required to return a small amount of the most relevant information. One challenge will be determining which information source best matches the customer's information needs.

In either case, the user interface will be critical. Even the most relevant information is worthless if the customer cannot understand the presentation (Saracevic and Kantor, 1997a). The best electronic information systems will have uniform but customizable, dynamic user interfaces that can smoothly integrate existing common data types (text, numeric, audio, video and image) from structured and unstructured sources with specialized types of data (maps, three-dimensional data, and continuous graphical data) and potentially new data types (Adam et al., 1996). These systems will incorporate algorithms and techniques that enable semantic interoperability, so that customers can
search in unfamiliar domains of knowledge (each with its own specialized vocabulary and ontology) using familiar vocabularies and ontologies (Atkins, 1997; Lynch and Garcia-Molina, 1995). There is an extensive literature on the problems of semantic interoperability and content-based retrieval.

Techniques must also be developed to effectively chunk or package information into units that do not overwhelm either the cognitive abilities of human customers or the physical capacity of the networks and systems transporting and storing the requested information. The rich human-computer interaction literature covers many of these topics.

Another important aspect of information location, is finding key relationships, especially in distributed, heterogeneous information sources. Data mining, the extraction of patterns, associations and anomalies from large data sources, is necessary from both the provider and consumer perspective. Providers are interested in access and purchase patterns, exploiting this information to improve decision making and identify potential customers. Typically customers are loath to provide this kind of information about themselves. Fraud detection, local vs. regional differences in customers and changes in customer requirements over time are other relationships and trends that providers are interested in. Customers are interested in identifying underlying trends in information, especially information that improves cost minimization and purchasing decisions. Although data mining techniques currently exist, they are not sophisticated enough to handle either the volume or the complexity of existing information. The extensive data mining literature explores the challenges in this area.

3.2 Providing Universal Access

The ultimate goal of many of national level initiatives (i.e., the National Information Infrastructure - NII initiative) with respect to electronic information, is universal access. This is consistent with the traditional goals of information providers, from an economic perspective and a social perspective. For universal access to be accomplished information providers need to solve the problems of integrating distributed heterogeneous information and information sources, designing and implementing effective user interfaces and solving the "vocabulary problem" (via semantic interoperability) (Schatz and Chen, 1996).

One of the challenges to providing universal access is devising techniques that will assist a wide variety of information display devices in handling voluminous, diverse and complex electronic information. Not only is there a variety of operating systems in the computer domain, but there will be a wide variety of other non-computer information display devices (e.g., palm tops, televisions, fax machines, video monitors, modems, and other information "appliances") to cope with. Accommodating legacy information display devices and receivers is probably a more difficult problem than accommodating and integrating legacy information and information sources (Tennant, 1997).

Another challenge is that there is a limited amount of bandwidth available for the transmission of electronic information that must accommodate an increasing number of information providers and customers and increasingly complex (and large) data sets. For equitable universal access to be achieved, intelligent use of the bandwidth, including the ability to guarantee bandwidth for a given period of time (in particular for law enforcement and emergency situations) must be identified and policies to support such uses enacted.
These challenges when combined with the economic pressures faced by information providers (in particular libraries) have led to a vision held by many of information providers behaving as “gate-keepers” and the service (or set of services) that they provide as a “gateway” to information or knowledge (Dowler, 1997; Drabenstott, 1993; Olsen, 1997; Rockwell, 1997). Indeed, several university libraries currently describe themselves in this manner.

4. Common Issues

As the use of electronic information increases, its impact on our lives also increases. Often this results in humans interacting with information and each other in novel ways usually without the benefit of policies, procedures and guidelines to govern appropriate behavior. Both information providers and customers come from a diverse and complex global community, each with their own unique blend of cultural perspectives. Furthermore, the roles of information providers and customers are blurring (Wiederhold, 1995). There are several social-economic issues, unrelated to the more technically oriented challenges discussed above that arise from this blend of diversity, universal access, novel interactions, and blurring of roles and boundaries. Some of the major areas of concern are: social issues, legal issues, quality and security issues, and economic issues (Lynch and Garcia-Molina, 1995).

4.1 Social Issues

There are several kinds of social issues faced by both electronic commerce and digital libraries. The major ones include:

- **Literacy** - in order to be an electronic information provider or consumer, a certain basic level of education or training is required, i.e., a basic competence in the operation of a computer or other piece of equipment that either generates or receives digitized information. Although some basic language and communication skills are required to interact electronically, multimedia frees producers and consumers of electronic information from the literacy levels currently required by printed information.

  A related issue is who will be responsible for providing these basic computer skills and training. Should training be freely available through public education systems or should it be part of the paid-for services provided by organizations in the information business? Will access to training as well as access to the appropriate equipment (computer) and facilities (an account and storage space on a server) separate society into information “haves” and “have-nots”? If so, what are the implications of this division?

- **Cultural Biases** - Filtering and organizing electronic information to assist customers in coping with the problem of information overload is a service which can be viewed as having the best interests of the customer in mind. However, there is also the possibility that the result, deliberate or unintentional, is that the cultural biases and social values of the service provider are being imposed on the customer (Atkinson, 1996). The simplest example of this is language bias. Should customers be required to access information in the language it was generated in, or should
part of an information service be the ability to translate information into the
customer's language of preference?

Translation of words (written or spoken) is relatively straightforward, but what
about translation of non-language information (i.e., images or music)?
Furthermore, information considered publicly appropriate for one group of people
may be offensive or even illegal for another (Kessler, 1996). One solution may be
to develop highly sensitive and individually customizable user-interfaces that could
accommodate a given individual or group's cultural and linguistic preferences
(Ferguson and Bunge, 1997).

- **Ethical Considerations** - The traditional librarian perspective is that information
  providers have a responsibility to ensure public access to information as equitably
  as possible socially and economically. This perspective has recently been
  augmented by marketplace forces in electronic commerce which also encourage
  information providers to support universal (public) access at a reasonable fee
  (Atkinson, 1996). However, not all governments, organizations or social groups
  support universal access, and may indeed actively attempt to restrict access to
certain kinds of information deemed inappropriate.

Universal access also encounters another set of ethical problems related to censorship
and cultural bias. As pointed out earlier, not all information is appropriate for all groups.
Different individuals and cultures have different opinions about the accessibility and even
the definition of material that could be considered due to background (racial, religious,
cultural), sex (including sexual preference), age and health including such information as:
pornography, material generated by hate groups and other racial or religious persecutors,
sexual predators (particularly child predators), drug dealers, terrorists and other criminals
(Lesk, 1997). Should there be some kind of a limit on who can be an information provider
(and if so, how could such a limitation be imposed and by whom)? Or should there be
limits on what kinds of information an individual customer can receive (who has the
authority to determine and impose such limitations, and how might it be accomplished)?

Now that almost anyone can be an electronic information provider, it is easier to
perpetuate other questionable ethical acts, for example, it is harder to detect plagiarism.
The sheer volume of electronic information makes it very difficult to enforce copyright
laws or to even detect illegal copies. False representation and false information can easily
be provided electronically, leading to concerns about information quality. These ethical
considerations are challenging enough within a given nation or culture, but electronic
information is available globally, and different nations and cultures have very different
perspectives, definitions, and social guidelines with respect to concepts such as plagiarism,
copyright laws, and “truth in advertising”. Do advances in information technology and
electronic information provision and consumption foster or encourage unethical conduct
or is ethical conduct indifferent to such advances (Rush, 1996)? How can internationally
accepted ethical codes be developed and enforced in light of these issues?

Privacy and privacy-related questions are another group of ethical issues that arise
with respect to electronic information. Providers want information about purchasing
habits, credit history, debt ratio, tax and other financial information, investment
preferences, employment history and other personal information that will help them better
target customers and identify new customers as well as provide the ability to determine if
price differentiation is possible between different customers. Often this involves information that individuals do not want organizations to have (and sell to each other). Related issues are concerns about who owns health and medical information and when can it be used to discriminate prices with respect to insurance policies and other health-related services or discriminate between individuals with respect to hiring, firing and promotion decisions. Similar concerns arise with respect to criminal record information.

- **Equality** - This set of issues involves questions such as, is there equal access to information and do individuals have an equally likely chance of providing electronic information. Experiences from some forms of electronic scholarly publication are very positive. A good example is the e-print archive for high-energy physics. In this case, access by status and country is more equal in the electronic version than the printed version as there is no need to be on any kind of distribution list (via social and professional connections) for electronic distribution (Ginsparg, 1997). The information is posted, and anyone can access it. There are other instances, in the vast biomedical collections for example, where the volume of information is so huge, that consumers tend to request information by a very small set of well-known and respected authors, journals, research centers, or some combination making it extremely difficult for newcomers to get recognized and accepted as information providers.

  This tendency also occurs in the marketplace. Consumers when faced with an overwhelming amount of information, tend to filter by “name-brand” companies or products, rewarding established organizations and making it extremely difficult for new enterprises to break into the market. Fears about the lack of quality control in electronic information drive this tendency even more dramatically. Information providers may have no difficulty making electronic information accessible, but how can consumers be encouraged to access the available information equitably?

  Another concern is that as the lines between the provision of information and information services become blurred between public entities and private entities, what incentives will there be to encourage private entities to provide their information and services freely and equitably to all? Won't it be in their economic interests to target certain information consumers and focus attention and accessibility on this group while ignoring others? What kinds of social mechanisms should be implemented to encourage equitable information availability?

- **Benefit** - Advances in information technology and information management clearly have dramatic societal implications (Bishop and Star, 1996). Education (Marchionini and Maurer, 1995), employment, the nature of work and the workplace, and the general quality of life are all impacted. It is not yet clear who will benefit from these advances and if the benefits will be equitable or not. It is reasonable to suggest that it may be important to investigate what factors influence the rate and degree of acceptance and adoption of electronic information over non-electronic forms and what social and economic mechanisms are effective in facilitating and moderating such changes (Bishop and Star, 1996; Ginsparg, 1997; Lynch and Garcia-Molina, 1995).
4.2 Economic Issues

Information providers need to get paid for providing information. Otherwise there is no incentive (in the private arena, for example a video store) or no ability (in the public arena, for example a traditional public library) to provide it. If a provider invests in equipment and other resources necessary to produce electronic information, does that give it the right to profit from the consumption of that information? How much profit is appropriate? If others modify the information (some kind of value-added service or simple modification), are they entitled to profit as well and if so, what is the equitable distribution of that profit (Rush, 1996)? According to Saracevic and Kantor (Saracevic and Kantor, 1997a) “the value of information rests with improvements in decision-making.” But the same information can be used in dramatically different ways with dramatically different results, so how can improvements in decision-making be evaluated, especially in the short term when payment is required?

In order to determine profit, an information provider needs to have some concept of the cost of providing the information or service. Unfortunately, the general consensus is that the current cost models and financial instruments used in traditional information production and consumption do not adequately address the needs of electronic information. Fixed cost models are insensitive to changes in content and costs. Electronic information comes in a variety of formats with different associated production and distribution costs. Cost models that are flexible and adaptable are required to handle the diversity and complexity inherent in electronic information (Adam et al., 1996; Choy et al., 1996).

A related value-based question is: As information services shift from labor-intensive to automatic and information technology-intensive processes, one of the questions becomes can a manual effort actually provide more value-added benefit (through greater precision for example) than an automated effort (Ginsparg, 1997)? Should there be a difference in pricing strategies for labor-intensive vs. automated information services? How will information providers determine the costs of providing previously unavailable information (Adam et al., 1996)? Historically, information provision and services are typically not broken down into individual transactions or monetized (Rockwell, 1997). As a result, not only do the institutions themselves have little idea what an information transaction is worth, but information consumers have no idea what the costs, appropriate price or value of the information is (Lesk, 1997). Indeed, this is such a difficult challenge that Saracevic and Kantor (Saracevic and Kantor, 1997b) developed a derived taxonomy to address the problem of determining or measuring the value of an information provider's (specifically a library) information and services. This problem is acerbated by electronic information which has almost a zero incremental reproduction and viewing cost, increasing the expectation that access should be free or extremely cheap.

Library services are typically subsidized, and the profession's ethic has traditionally been that services should be offered at no cost to the public. Therefore, librarians are typically more concerned about the possibility that economic inequities would lead to some consumers opting not to seek information thereby creating a class of informationally and economically poor than determining prices and costs for individual service transactions. As a result, cost containment has been the major method of trying to maintain an acceptable level of service in the face of increasing expenses. Technological
advances such as client-server architecture and international standards for software and documentation (i.e., SGML), the use of Commercial-of-the-shelf (COTS) products and reusing or sharing resources among institutions were the primary cost containment techniques (Drabenstott, 1993; Rockwell, 1997).

Nonetheless, economic information is not free. Some method of compensation is necessary. Currently, there are at least two basic electronic information compensation models: 1) allowing free access but charging for content (i.e., freely accessing the index and table of contents, but charging for anything more) and 2) charging for access but allowing free perusal and consumption of the content (Lesk, 1997). These two models are not mutually exclusive and conflicting, as both comfortably co-exist in cyberspace.

Many different electronic information funding models have been proposed, but the basic models are either time-based (payment for unlimited access for a given unit of time, e.g. a month), request-based (payment per request), or some combination of both (Adam et al., 1996; Drabenstott, 1993). Some possible models include (Lesk, 1997):

- Institutional (public and private) subsidies - the current model for most information providers;
- Free general services and charging by transaction for unusual services, especially those requiring any human intervention - an existing viable business model (i.e., warranties) that may make sense for all information providers in the future;
- Charging consumers for everything - assumes that producers can cost and transactionalize information services (for example this model works in video stores). Some common suggestions include charging by: connect time, CPU usage, fee-per-search, fee-per-hit or retrieval, and download fees. A problem with all of these suggestions is that in general consumers, do not understand how they work. This creates a situation where charges appear to be unpredictable, resulting in unreasonable or unpredictable consumer behavior which is distressing to information providers;
- Subsidizing services through providing advertising space (typical of magazines, television, and WWW);
- Other subsidizing mechanisms perhaps including pledge breaks (public appeals for donations similar to public television and radio);
- Taxes or other sources of public funds;
- Subscriptions (pay for a given length of time i.e., a year) or licenses (viable concept in the software market);
- Memberships similar to “buying clubs” where individual consumers pool their resources to allow access to information (pricing issues could be resolved via price discrimination, non-linear pricing and service bundling (Kluisters, 1997). This model and the subscription model could include “bounties” for signing up new members;
- Charge information providers a per-unit fee for the “privilege” of having their information and services accessible, then charge information consumers for the nominal incremental cost of accessing the information (similar to a per-page author charge under consideration by some journals);
- Opportunity cost - measures the opportunity cost of providing information or a service as opposed to measuring the cost of expended resources. “Opportunity
cost is determined by the relationship between supply and demand for a given resource, so that the opportunity cost of an idle resource is close to zero but that of an overutilized resource is so high that it is basically unaffordable” (Adam et al., 1996); and

- Using a detailed byte-by-byte charging algorithm - an interesting idea from Ted Nelson (CNRI) and CMU’s NetBill project (Lesk, 1997).

Electronic information economic models require a series of specialized costing and pricing algorithms that can dynamically determine the cost and price of an information or service and modify the model with a variety of environmental factors including supply and demand (Choy et al., 1996). From an information consumer's perspective, these algorithms need to provide cost minimization and multiple provider billing (Paepcke et al., 1996). From an information provider's perspective, the algorithms need to rapidly and dynamically respond to changes in supply and demand while exploiting new marketing opportunities (Lesk, 1997).

Some interesting electronic financial instruments have been developed extending the concept of electronic information to payment methods. They include:

- Digital or Electronic Cash - turns real currency from a banking institution into digital cash (“cyberbucks”). Security is handled via public key encryption. An example is DigiCash's Ecash (http://digicash.support.nl).
- Electronic Wallet - electronic transactions are charged to a credit card account. Security is handled via public key encryption. An example is CyberCash Wallet (http://www.cybercash.com/).
- Electronic Data Interchange - EDI - trading partners agree to exchange transaction information directly. Typically direct access is provided via dial-up connections or a proprietary network.

4.3 Legal Issues

There are several issues related to laws and governmental policies. Since electronic information exchange occurs on a global level, national governments will have to negotiate an international level policy framework that can accommodate the exchange of information across international boundaries and differences in cultural values and laws (especially with respect to copyright, intellectual property, privacy, information ownership, fraud and other business crimes, taxation and currency exchange) (Prentice, 1997). Just as an example, the current export regulations on encryption systems (presumably for reasons of national security) significantly inhibit the development and implementation of a secure, worldwide network infrastructure (Adam et al., 1996; Lesk, 1997).

Other more localized legal issues surrounding electronic information include (Lesk, 1997):

- Unauthorized access - electronic information appears to be more vulnerable to unauthorized access, theft and fraud than physical copies as it is harder to detect. A variety of techniques are being investigated to help insures the security of electronic information, including such topics as “firewalls”, electronic signatures, encryption, special “rendering” or viewing software or hardware, and electronic watermarks.
• **Liability** - traditionally US law distinguishes between authors and publishers who were held responsible or liable for information that they produced and organizations like the post office, libraries and bookstores who merely distributed the information. Electronic information providers can distribute as well as produce information. Many system administrators are aware of the difficult legal questions regarding their responsibility and the responsibility of the organizations they work for (especially public ones such as universities, or private subscriber on-line service providers like AOL) for information published, displayed or distributed from their sites.

In situations where electronic information has multiple authors, and multiple versions how can expertise be determined and liability assigned? For example, according to current US law, the publisher or author of a book that contains bad advice on investments is not liable yet a stockbroker is liable for bad advice on investments. How will electronic information be handled? Who will determine who is liable, when, and for what?

• **Trademark Infringement** - in the US two organizations can have the same trademark or name as long as they are well-distinguishable and separable businesses. Commonly used examples are: HP hot sauce and HP electronics (Hewlett-Packard) or Sun Oil (Sunoco) and Sun Microsystems. With the current Internet addressing systems, only one of them can have the .com address (for example, Sun Microsystems owns the www.sun.com address). Savvy Internet individuals realized this early in electronic commerce and registered for addresses that contained trademark names of large or popular organizations. Consumers guessing at the Internet address of an organization, product or service based on its trademarked name, may or may not connect to the appropriate site.

Another related problem is that individuals can copy or scan in trade-marked images (for example a state or university seal or commercial caricature), and use them as wall-paper or images on their own personal web pages. Similar to copyright protection, many organizations require notification and/or payment for permission to use their trademarks as it is often interpreted as an organizational endorsement.

• **Copyright and Intellectual Property Rights** - copyright issues, in particular copyright violation and the related intellectual property rights issues are probably the major legal concerns with respect to electronic information. It is generally agreed that without some form of intellectual property protection and reward system, many information providers will have no incentive to generate the information, at least not in an electronic form (Lesk, 1997). Pamela Samuelson (Samuelson, 1995) from Berkeley is a well-know authority on the topic from the perspective of electronic information.

Virtually anything that can be copyrighted, can also be digitized. Once digitized, anyone with a computer can copy it, modify it and distribute to anyone else that has access to a network. The regulations that exist today - which include: no downloading at all; no electronic storage (view only); no copies or distribution, even internally; no copies or distribution to third parties; and specific limitations on various types of use - are largely
ignored by information consumers (Kalakota and Whinston, 1996). Information providers implicitly endorse this behavior by “looking the other way” in many instances.

Changes in US laws with respect to the definitions and legal treatment of electronic information, especially its transmission are being proposed. These changes are not necessarily in the best interests of consumers. For example, if digital transmission is no longer defined as “publication” but rather as “copying”, then providers of the information are not required to file a copy with the Library of Congress, and are not subject to the current copying rights available under the “first sale” and “fair use” doctrines. Legal rights with regard to copies are more restrictive than publishing legal rights (Lesk, 1997).

Most experts are of the opinion that new copyright laws and practices, at least with respect to electronic information are going to have to be created as the speed of technological advancements have left the legal systems far behind (Kalakota and Whinston, 1996; Unsworth, 1997). Electronic information is easy to copy and redistribute, but it is difficult to detect a valid copy from an illegal one. New copyright laws and practices must be enforceable and therefore will probably rely on new technology to help protect copyrighted material from unauthorized access (Ching et al., 1996; Garrett and Lyons, 1993), reproduction, manipulation, distribution, and performance or display. New technology will probably also assist in the detection of copyright violations through new methods of authentication, management of copyright protected material (such as the clearinghouse model used predominantly by the music industry), and licensing techniques (Kalakota and Whinston, 1996).

Copyright laws and laws regarding intellectual property rights, key to the viability of certain types of electronic information generation, vary dramatically across international boundaries. Attitudes, biases, definitions and values with respect to these issues also tend to vary across cultures within national boundaries (Mansell, 1996).

There are several methods currently under investigation to protect intellectual property rights and copyright of electronic information. They include: (Lesk, 1997)

1. Fractional Access - this works only for very large information sources (e.g., LEXIS/NEXIS) as the value of the information source lies in the volume of information and the knowledge that can be gleaned from analyzing or “mining” the entire information source. There is no economic advantage to copying small portions of the data, and illegally copying the entire data source should be relatively easy to detect.

2. Interface Control - this solution requires a proprietary interface, implying that universal access is no longer possible as only information providers and consumers with access to the proprietary interface can produce and access the information.

3. Hardware Locks or “dongles” - this is the hardware equivalent of interface control (a software solution). Access to information is restricted to those individuals who have the proprietary access hardware (video games such as Sega or Nintendo are good examples). Furthermore, this solution has encountered consumer resistance as it is expensive and requires frequent upgrades typically not compatible with older software.

4. Information Repositories - in this solution legitimate copies are only available from one large repository or source. Any other copy is not legitimate. Some organizations exploring this approach include: InterTrust (previously EPR - Electronic Publishing Resources) and CNRI (Corporation for National Research Initiatives, currently working
with the US Copyright Office) both from the US and Imprimatur (Intellectual Multimedia Property Rights Model and Terminology for Universal Reference) from Europe.

5. Steganography - this solution involves the embedding of hidden messages in the information. Each legal copy is labeled with a different identification number and illegal copies could thus be tracked back to the original purchaser (i.e., “digital water-marks”). The major problems are that the “hidden” codes or messages are easy to remove and hard to insert and while it appears to work with complex images it does not work with simple text and may not even apply to audio data.

6. Encryption - in this solution the information is encrypted (sometimes in cryptolopes or secret envelopes), and cannot be interpreted without the encryption key (software or hardware dependent).

7. Economic Approaches - these solutions attempt to identify ways to make it uneconomical to pirate or illegally copy electronic information. Some examples of ideas include: provider page charges to reduce the per-copy price, site licenses to reduce on-site cheating, and advertiser supported publications.

8. Flickering or “Wobbling” - these solutions employ information technology that allows an information consumer to view but not capture information (Lesk, 1996).

4.4 Quality and Security Issues

There is a definite need to build credibility for on-line information. Many information consumers are of the opinion that “Information that you get for nothing over the Internet is worth nothing” (Lesk, 1997). This opinion is founded primarily on the well-know low quality of information on many netnews groups, and the lack of information quality control on personal home pages. Unfortunately none of the existing searching engines have a way of evaluating electronic information quality (traditional information quality cues are typically missing) and therefore no way of sorting or filtering information by quality.

This lack of information about quality tends to encourage information consumers to limit searching to known experts (authors, WWW sites, organizations, publishers or journals, etc.) or information recommended by known experts. Some search engines do allow consumers to profile an expert or group of experts and request similar information (i.e., give me information identical to what Joe Einstein requests). Information integrity is still a problem however because electronic information is so easily modified. An information consumer is rarely guaranteed that the information of interest was truly generated or endorsed by the expert.

Currently, electronic scholarly publishing is a hotly debated topic. Much of the concern centers around the questions of the information and intellectual quality (Drabenstott, 1993; Peek and Newby, 1996). One of the quality-related concerns that can be addressed electronically is peer-review (a traditional form of quality control in academic publications). However, without peer review mechanisms, the quality, validity and viability of the electronic information in is at best, suspect.

Ginsparg’s observations about the on-line e-print system (high particle physics) support the belief that high quality, valid, electronic scholarly communication and publishing is possible. “The electronic form, once posted to an archive is instantly publicized to thousands of people, so the embarrassment over incorrect results and
consequent barriers to distribution, is if anything increased (not decreased). Such submissions cannot be removed but can be replaced by a note that the work has been withdrawn as incorrect, leaving a more permanent blemish than a hard copy of limited distribution that is soon forgotten“ (Ginsparg, 1997).

Issues of security and control are related to quality issues. Electronic information needs to address security in at least four areas (Adam et al., 1996; Kalakota and Whinston, 1996):

- **confidentiality** - protecting access to the contents of electronic information (especially sensitive electronic information, such as personal financial or health information, and strategic business or national information) from unauthorized access and distribution.

- **authenticity** - not only attributing electronic information to the correct information provider, but validating the information as original, accurate, and correctly attributed. This can be especially difficult in multi-authored and multi-versioned electronic information environment (Wiederhold, 1995). One area where authenticity is important, but it is equally important to preserve confidentiality is in electronic peer review (or a validation process of any kind where anonymity is important).

- **integrity** - protecting the contents of electronic information from unauthorized modification. This type of security involves a balance between easily enabling authorized updates and preventing unauthorized ones. The authenticity of modifications must be verifiable (also challenging in a multi-authorized and multi-versioned electronic information environment).

- **privacy** - protecting the access and usage patterns of information consumers from unauthorized access and resale. This form of security was discussed earlier.

An important challenge for the implementation of any security technique is to balance the need for security with the need for performance (access and timeliness). Authorized access and modification must not be so difficult that it is never attempted, or abandoned before completion. Information is not valuable if it is not accessible, timely or useful. Likewise the validation techniques while they must be as accurate as technically feasible, cannot be so time and resource intensive that the accessibility and timeliness of the information is compromised.

### 4.5 Standards

There is an especially important need for internationally accepted technical standards with respect to the representation, formatting, transmission, and protocols for electronic information. This is the only way to ensure compatibility and therefore interoperability between equipment, data, practices and procedures, all of which is necessary to achieve the goal of universal access (Kalakota and Whinston, 1996; Kessler, 1996; Paepcke et al., 1996; Tennant, 1997) and global electronic information exchange. Yet as pointed out earlier, there are many social, cultural and political barriers to overcome when developing usable and acceptable international standards, even when the benefit to all is clear.

There are several international organizations concerned with standards. The International Organization for Standardization (ISO) is involved in many different kinds of standards development, including standards that relate to electronic information (e.g.,
SGML). Another is the Internet Engineering Task Force (IETF) - see http://www.ietf.org - which is specifically interested in Internet architecture and smooth Internet interaction and operation (Kessler, 1996).

At a national level, while document standards such as SGML, HTML, TEI (Text Encoding Initiative), VRML (Virtual Reality Markup Language), and MARC (Machine-readable Cataloging) exist, in practice much of the electronic information interaction occurs via e-mail, anonymous ftp, Gopher, and WWW browser platforms with TeX, LaTeX, PostScript, PDF, ASCII text, Word and WordPerfect formatted documents. Most of these formats do not have existing mechanisms to distinguish the contributions of multiple authors and multiple versions, nor do they have the ability to include active links to other electronic information. Many of the formats used in practice are commercial, and therefore proprietary, which means that they are not platform independent and cannot be readily transmitted or accessed by all. Will common practice dictate what standards become accepted or will some governing body take responsibility for thoughtful and independent (unbiased) planning and design? If a set of standards is accepted and adopted, what kind of translation capability from these “legacy” formats will be provided?

5. Conclusions

There are many challenges and important issues associated with electronic information that have not yet been solved. Indeed they are in part responsible for “slowing down” the growth and expansion of electronic information in our everyday lives. But the advantages, especially the economic ones, appear to outweigh the disadvantages and most will agree that electronic information already has a major impact on our lives which will only increase as advances in information technology and information management continue.

While it is not known what the future will look like and what solutions will be brought to bear on many of the challenges and issues discussed here, one thing is known for sure. Whatever solutions are developed and implemented will never be final. Electronic information and the environment that it has created will continue to evolve. New challenges and new issues demanding new approaches will appear as new technologies become available. As electronic information changes and evolves the needs of information providers and consumers will change and evolve, perhaps in ways that we cannot yet imagine. The result is a self-perpetuating cycle in electronic information of new technology creating new challenges, situations and needs which in turn will drive the investigation, development and deployment of more new technology.
References


Atkins, D. E., “Report of the Santa Fe planning workshop on distributed knowledge work environments: Digital libraries.” Supported by a Grant from the National Science Foundation (NSF-IRI-9712586) to the University of Michigan School of Information, March 9-11 1997.


World Wide Web Sites

DigiCash's Ecash, http://digicash.support.nl
Internet Engineering Task Force (IETF), http://www.ietf.org
Digital Library Initiatives
University of California at Berkley, http://elib.cs.berkeley.edu/
University of California at Santa Barbara, http://alexandria.sdc.ucsb.edu/
University of Illinois at Urbana-Champaign, http://dli.grainger.uiuc.edu/
University of Michigan, http://www.si.umich.edu/UMDL/
Stanford University, http://walrus.stanford.edu/diglib/
Index Terms

Access value
Collaboration
Compensation models
Content-based search
Copyright
Cost models
Cultural bias
Economic models
Economics
Electronic Information
Ethics
Filtering
Gateways
Global networks
Inflation
Information acquisition
Information identification
Information location
Information ownership
Information providers
Information retrieval
Information seekers
Information storage
Intellectual property
Multimedia
Preservation
Privacy
Quality control
Scholarly communication
Security
Semantic Interoperability
Standards
Universal access