Electronic Commerce and Digital Libraries: Towards a Digital Agora*

Editors: Nabil Adam and Yelena Yesha

August 17, 1996

1 Introduction

Electronic commerce (EC) and digital libraries (DL) are two increasingly important areas of computer and information sciences. EC is concerned with creating systems that carry out business interactions and transactions that take place on behalf of consumers, sellers, intermediaries, and producers. DL on th e other hand is concerned with creating systems that carry out interactions among information and knowledge producers, librarians, and information and knowledge seekers.

Both EC and DL share the goal of providing the ability on a global scale to acquire, store, and retrieve information electronically. An additional goal of EC is to facilitate business transactions.

EC/DL systems are typically a collection of a distributed autonomous sites that work together to give the user the appearance of a single cohesive collection. They deal with large amount of multimedia information where content may be stored on variety of computer media such as disk, tape or CD-ROM and typically comes from a variety of sources who may wish to control its use (retrieval or modification) or to add value to the content. Users of these systems have a wide variety of domain and computer expertise.

*Contributors: Nabil R. Adam, Rutgers University; Baruch Awerbuch, Johns Hopkins Univ.; Karen Bennet, IBM; Barbara T. Blaustein, The MITRE Corporation; Alex Brodsky, GMU; Ray Chen, UMBC; Bob Grossman, Univ. of Illinois; Jim Johnson, Eikon Strategies; Kostas Kalpakis, UMBC; Catherine D. McCollum, The MITRE Corporation; Anna-Lena Neches,USC/ISI; Bob Neches, DARPA; Arnon Rosenthal, The MITRE Corporation; Jacob Slonim, IBM; Howard Wactlar, CMU; Ouri Wolfson, Univ. of Illinois; Yaacov Yesha, UMBC; Yelena Yesha, CESDIS/UMBC

2 Problem areas

One may divide EC/DL research into the following problem areas:

1. *Acquiring information.* There are two sides to this problem. First, we need to deal with the problem of digitizing currently available information such as books, films, maps, audio, etc. In addition to the technical aspect of this problem there is also the economic aspect (costs of digitizing the information). Second, as new information is made available we need to have mechanisms to capture and store this information.

2. *Storing information.* An EC/DL's storage system must be capable of storing a large amount of data in a variety of formats and accessing this data as quickly as possible.

3. *Finding information.* In an EC/DL system, consumers will need on-line facilities that not only help them in retrieving information and locating resources but also provide help in retrieving information and locating resources that match certain expectations and desires. For example, in an DL system a user may be interested in locating an image of the Hudson river with some buildings nearby. In an EC system, a consumer may be interested in finding products within certain price ranges. Matchmaking programs will bring together producers and consumers. Examples of such a service include:

   - Mechanisms to describe information services by capability rather than by name.
   - Mechanisms to perform content-based search.
   - Methods to integrate various search engines.

4. *Using the information.* Coupled with information repositories, an EC/DL system must also allow consumers/users to act on the information that the system contains. For example, consumers must be able to complete transactions on products and services if they so desire and if the providers support such a model of interactivity. User interface is another important aspect of this problem.

5. *Securing information.* On one hand the system must be secure against malicious use, misuse and data-corruption failure and must ensure that

the privacy of its users as well as the intellectual property of the vendors and information producer are protected. On the other hand, however, the system should provide open access so that vendors and information producer can add/update information and services any time. Authentication and security services such as privacy, authentication, integrity, confidentiality, authorization, and controls need also to be dealt with.

6. *Universal access of information.* In an EC/DL environment today, there is a multiplicity of clients. Users may use TV sets, radios, PCs, PDAs, laptops, and cellular phones for accessing information. The goal of the National Information Infrastructure (NII) project is to allow universal access to distributed stores of information (Digital Libraries) and to provide this access at a reasonable cost to every citizen [?]. Widespread use of EC/DL is predicated on the universal availability of access for a large number of users [?]. EC/DL systems will have to contend with many different information appliances. In such a heterogeneous world of objects and clients, the issue of Interoperability assumes paramount importance.

7. *Socio-economic impact.* Public policies, regulations, laws and just plain public opinion can stop the development of technology in its tracks. We need to explore ways to address many of the social and economic issues either by new technology or modified technology. We need to create technological answers to many of the public policies, regulations, laws, social/cultural, and economic issues. For these technologies to be effective, there is a need to first be aware of such problems and concerns and second to establish dialogue with researchers from different disciplines.

3 Case Studies

3.1 Electronic Commerce

3.2 Digital Libraries

In September of 1994, the joint NSF/ARPA/NASA Digital Library Initiative awarded grants to 6 universities and industry consortium headed by Carnegie Mellon University, University of California, Berkeley, the University of Michigan, the University of Illinois, the University of California, Santa

Barbara, and Stanford University.

The focus of the Carnegie Mellon project is the storage and retrieval of video materials for science and mathematics subjects and targets users in the K-12 educational community. The project combines speech recognition, image understanding and natural language understanding to both catalog video and to retrieve relevant video segments. It views a digital library from a video library perspective. e.g. that video is the primary content and that additional contents of the DL such as text and still images are derived from the video source. This viewpoint is justified by the reasoning that video is the most difficult of data types to be stored, searched and manipulated due to its fixed temporal nature. Users have little control over the rate at which information is presented to them. A digital video library, therefore should allow users the freedom to browse video in a variety of manners such that they may take in or skip over information at will. Continuous speech recognition is us! ed to transcribe narrative and dialog from video into on-line text. Natural language processing is employed to achieve a natural language user interface in both spoken and typewritten form and to process video transcripts for storage and searching. Security and billing mechanisms are required to uphold copyright and facilitate payment for viewing videos. Research issues include authenticating and protecting the privacy of individuals and support for various pricing models and the user's reaction or behavior when faced with various pricing models.

The focus of the U.C. Berkeley project is to develop digital library technologies to manage a wide variety of environmental data for the state of California including reports, computer models, maps, plans, aerial and ground photographs, videos and structured databases. Users representing a broad range of prior subject knowledge will access the digital library. The U.C. Berkeley DL is centered on distributed object oriented database technology. Their overall approach is to adapt database technologies to meet the demands of digital libraries. Other components of the digital library include powerful interfaces that allow users to manipulate multimedia documents (not simply retrieve and view) and special network protocols adapted for distributed information retrieval.

The focus of the University of Michigan digital libraries (UMDL) project is on making earth and space science materials available to a wide range of users via a single interface. Agent technology is a key component to achieving this goal. Content for UMDL presently comes from existing collections of science journals. Additional efforts include storing and accessing spatial data such as social science and earth science data. A digital library is viewed as a distributed collections of digital documents each represented by a different query interface and storage and retrieval mechanisms. UMDL differs from other DL projects in that the goal is not to bring all documents into a single system. Software agents are used as a means of carrying out a user's requests in an open market fashion on these heterogeneous collections. The main ares of research in UMDL are agent based technologies and user interface issues. Software agents play a central role in the UMDL system. User Interface Ai gents (Remora agents) are employed by users to periodically check on the contents of a digital library and report any changes to interested users. Collection Interface Agents perform queries on specific document collections. Thus a given CIA must know the specific query capabilities of a collection. Remora agents follow a market based approach in that they sell their services to users and buy the rights from collection interface agents to make searches on various collections. Auction agents can mediate the buying and selling process by making potential buyers and sellers aware of one another. Agents communicate with one another using an enhanced version of KQML. These enhancements include the ability for an agent to register itself and to advertise its capabilities.

The focus of the University of Illinois at Urbana-Champaign digital libraries project is to facilitate access to engineering and science journals. Articles are obtained directly from the publishers in SGML format and are made available for searching and full display including text, figures, tables and equations. Research is generally aimed at the scalability (hundreds of thousands of documents with thousands of users) and functionality of digital libraries. Building effective document indexes is also a prime initiative. The UIUC effort, however, treats heterogeneous collections of documents as a federated system with a central repository for indexes and a search interface consisting of multiple views of the source collections. Users should be able to query across collections in a consistent fashion and retrieve and view documents in their entirety.

The focus of the University of California - Santa Barbara project is on the management and retrieval of geographic and spatially indexed information. This approach differs from other DL projects in that short term goals focus on developing a rapid prototype system using existing components as a basis for comparison with future research. Within the first year of the project, a prototype system was constructed using commercial database management systems (Sybase) and geographic information systems (ArcView). This prototype was then augmented with a WWW interface that has recently been deployed for testing. The content for the prototype systems include aerial photographs from NASA/AMES Earth Resource Air Photo Library and air photos over time of the California and Santa Barbara areas. Additional holdings include materials from the Sierra Nevada Ecosystem Project (SNEP) which consists of physical, biological, cultural and political features from surrounding areas stored as ARC/INFO coverage.

The focus of the Stanford digital libraries project is to tie together a broad range of information sources from traditional stores of document collections to personal information stores. All resources are to be made accessible under an information bus that will tie together users, services and resources. Content will be targeted at computer related literature and will be provided by existing sources such as the SNRI Computer Science Technical Reports collection, the Stanford University Folio System, public domain materials and materials from the participating publishing companies. The user community will include faculty, staff and students at Stanford, scientists at NASA's Ames research center and customers of the various publishing companies. The research in the Stanford project centers on the information bus, a conceptual architecture that encompasses low level networking protocols, agent based communications protocols, information sources and information clients. The information bus facilitates the exchange of information on a variety of levels and is used to support digital library services including searching, translation, publishing, authentication, copy detection, financial, mediator and personal information services. The initial research for the project is concerned with developing the specification for an information bus that can accommodate the variety of modalities present in a digital library.

4 Challenges

The goal of this section is to examine the challenges in providing efficient and effective EC/DL systems and to outline some research directions.

4.1 Acquiring information

The vast majority of potential EC/DL content is not currently in digital form. Millions of books, articles, films, audio recordings, maps and other sources are presently stored in their physical, non-digitized form. Bringing such artifacts online is not trivial and many researchers have addressed the basic problem of digitizing existing media.

Documents such as books, magazines, newspapers, journals and correspondence make up the majority of the content one finds in traditional libraries. Consequently, most of the research in digitizing has focused on scanning and storing paper documents. Since this operation is not trivial (i.e. it costs money), it warrants further investigation into the mechanics involved.

Comparisons have been made between traditional library storage of books and digital library storage of books. In [?], the authors contend that storage and maintenance costs for an average textbook in a DL and in a traditional library are roughly the same. An average bound book can be scanned for about $50, with the bulk of the cost going to the labor involved in scanning each individual page [?]. If the book can be fed to the scanner using an automatic page feeder, costs drop to around $10 per book [?]. High speed drum scanners are currently able to scan a single page per second.

Given that it is possible to scan various documents, the next question arises as to what resolution to scan and store the page images. In general, higher resolution scans pick up more detail but also produce data files that are larger.

4.2 Storing information

4.2.1 Storage

Text-only documents stored in formats such as ASCII, LaTex, HTML, SGML, and PostScript are by far the easiest to store. Digital audio and video are more difficult to store because they require significantly more storage space and their delivery is time-dependent.

Compression techniques save storage. For text-only documents, the Unix compress or freeware gzip utilities provide anywhere from 10- to 60-percent compression. Several compression standards exist for digital images (JPEG), audio (uLaw), and video (MPEG). EC/DL collections that are too large to store entirely on a disk use hierarchical storage mechanisms. In an HSM, the most frequently used data is kept on fast disks while less frequently used data is kept in near-line such as an automated (for example, robotic) tape library. Using data-usage statistics, the HSM can automatically migrate data from tape (near-line) storage to disk (on-line) and back, as required.

To meet the increasing data storage needs, new storage technologies utilizing three dimensional volume storage are emerging to replace the current two dimensional storage systems. In particular, the 3D optical holographic storage offers promise as a high capacity and high throughput storage technology. In a holographic storage system, the data are stored in page form in a distributed holographic pattern. Therefore it is less sensitive to localized defects on the storage medium. This build-in redundancy further enhances the speed and reliability of the data storage system.

4.2.2 Classification and indexing

. Classifying and indexing objects is filled with pitfalls, however, because individual perceptions vary. Another complicating factor in indexing and classifying is the tremendous amount of potential content that remains to be indexed. It is clear that manual methods for classification are insufficient for all but the most trivial EC/DL system.

Automated classification systems differ significantly in their approaches, depending on the type of content under consideration. Classifying short stories is quite different from classifying maps, both in terms of the mechanics involved and the appropriate classes. These distinctions make current automated classification efforts highly domain-specific. Image classification approaches are conceptually different from those used for text classification. Although many domain-specific systems allow "content-based" querying, most are relegated to a very narrow range of images and may require the services of human classifiers. Video classification and indexing requires systems that can parse video into manageable portions, typically called camera shots. As with image classification, the type of classification and indexing performed on video is driven by the types of queries posed by users. The classification of audio, musical notation, and maps presents additional research challenges. Certainly it is possible to videotape such artifacts from a number of angles and store this within the digital library. However, for many items, this would not afford the interaction necessary for some scholarly research. In the foreseeable future, it will therefore be necessary to incorporate these artifacts within the scope of EC/DL.

4.3 Finding Information

In the EC/DL domain, there are a variety of information-retrieval techniques, including metadata searching, full-text document searching, and content searching for other data types. The success of information retrieval can be measured in terms of the percentage of relevant and extraneous information retrieved. Techniques to improve retrieval effectiveness include preprocessing documents to extract additional metadata before storing them in a document database. Several researchers are focusing on automating the creation and maintenance of user profiles and applying these profiles to information retrieval. Software agents are an extension of filtering techniques, although filtering tends to imply passive mechanisms whereas the use of agents implies a more proactive approach. Software agents also embody the notion of improving over time as they record additional user actions and reactions.

4.3.1 Feature Extraction

We consider the problem of identifying and efficiently extracting effective and robust features from multimedia data as a problem whose solution will enable highly sophisticated functions. This includes extracting features from the whole range of data in EC/DL systems, text, audio, images, and video. Feature selection and extraction provides the metadata to multimedia data manipulated in EC/DL systems. However, the queries over those metadata are not limited to traditional exact match queries, but they include spatial, temporal, and other logical relationships (usually domain dependent), while providing for approximate matches [8, 6, 20, 5].

Usually, automatic or semi-automatic feature extraction from multimedia data is limited to a projection of the data in a space spawned by rather primitive atoms, and to aggregate functions over such projections. In the case of text, common features include frequency counts of words, phrases, and n-grams. In the case of images, common features include, among others, color histograms, edge statistics, boundary representations, statistics gathered from a segmentation of the image into regions (e.g. rotation, translation, scaling, and reflective invariant moments), texture, and statistics of image interest points (e.g. points of change of curvature, corners, etc) [8, 11, 13, 6]. These low-level features can be used to construct higher-level (logical) features of the image. To this end methods from model-based object recognition and syntactic pattern recognition can be employed. Unfortunately, these methods require a high degree of human intervention and are limited to small application-specific domains. Knowledge acquisition techniques seem promising in assisting or even semi-automating the extraction of such higher-level features from images. In the case of video, the standard approach is to partition the video into contiguous sequences of frames, construct a representative frame from each such sequence, and extract features, such as those mentioned earlier, from each frame (treating it as a still image). In addition, motion sensitive features are extracted from the sequence of representative frames [6, 5]. The situation is similar in the case of audio, however different features are sought (e.g. spectrograms, phonetic codebooks).

Event hough there is a very large body of work in image and speech processing, image analysis, and pattern recognition, addressing feature extraction issues, and incorporating them into a EC/DL system is far from being adequately addressed. The search for efficient computable, effective, and robust features, both global and local with respect to image, video or audio clip, is far from over. Furthermore, considering that a multimedia object may contain text, still images, video, and audio segments, the task of utilizing the features for each individual segment to characterize that object for retrieval purposes is both challenging and promising. In other words, the question is to select and extract features from a multimedia object that can be used to retrieve similar (approximate equivalent) multimedia data objects.

There is an interplay between features and similarity, in that to evaluate the effectiveness of a feature one needs to have a working definition of similarity (or approximate equivalence) that is as close to the human notion of similarity as possible. The usual approach is to select a set of transformations and a distance function over sequences of such transformations, and then one declares two entities A and B as being approximately equivalent if one can convert, using some of the selected transformations, entity A to an entity whose distance from B is no more than a user selected threshold. Examples of such transformations and distance functions include various norms in vector spaces, edit-distance in strings, and set-theoretic functionals. However, the question of which set of transformations and which distance functionals capture best the human perceived similarity is not adequately addressed [3, 14]. Moreover, the question of designing adaptive and learning distance functionals, that adapt and learn from the user's feedback, needs further investigation [9].

Assume that a suitable set of features has been selected, and automatic extraction methods are available. Designing a query language and indexing schemes to process queries efficiently is the next important step [8, 6]. The query language is highly influenced by the notion of approximate similarity one adopted in selecting the features. For example, while point-access methods are well studied, are suitable for indexing numerical features. New indexing schemes are needed for non-numerical features such as graph-based representations. The vast available experiences in databases, knowledge representation, programming languages, and computational geometry, and other related areas, need to explored and customized in the application areas of DL and ECs.

Another consideration is the perceptual measurement of image quality. It is well known that the widely used mean squared error measure is not adequate for assessing the quality of images as perceived by humans. A quality measure that can be efficiently computed and at the same time correlates well with the opinions of human viewers is needed for the design and evaluation of image processing systems such as image compression systems. Such perceptual measures, including measures based on models of the human visual system, have been introduced by researchers. One example of such work is a paper by Heeger and Teo [12]. We believe that, due to the complexity of the human visual system, there is a need for additional interdisciplinary research that will involve researchers in the areas of image processing, optics, neurobiology, and psychology, and will result in improved models of the human visual system that will enable the introduction of improved perceptual image quality measures.

4.3.2 Data Mining

The digital agora is the nexus of the emerging infrastructure to support electronic commerce, digital libraries. generation applications requiring the searching, browsing, ordering, retrieving, and paying for both digital and non-digital goods and services. To date, the majority of the focus has been for the development of technology for the direct support of this infrastructure: how can information be retrieved by attribute and keyword, how can electronic catalogs be developed, what are the appropriate payment mechanisms, and how can the transactions in the agora be made secure?

As the use of the digital agora grows, there will be increasing amount of information collected, the most important of which will be 1) the access patterns to information in the agora by patrons and 2) the purchase patterns of goods and services in the agora by customers. A fundamental challenge is to develop algorithms and software for the indirect support of the agora by understanding the information in these patterns and exploiting this information as a basis for better decision making.

Here are some examples:

- Given the access patterns to information in the agora by patrons, we can ask: What information is accessed and by whom? Given a new document, how likely is it that it will be accessed and by whom? What documents are likely to be accessed together? Which subjects are becoming more important as indicated by trends in the number of documents available and the number of requests for them? Which documents are naturally grouped together on the basis of patrons requests, or sequences of requests?

- Given the purchases patterns of goods and services by customers in the agora, we can ask: What goods and services are purchased and by whom? Given a new product, how much of it is likely to be sold and to whom? Are there local and regional variations in the purchases of goods and services? Which transactions are fraudulent? How are the purchases of a customer likely to change over time? Uncovering these types of patterns, associations and anomalies in transactional and operational data is becoming known as data mining. There a variety of techniques that have been developed for this, including the use of classification and regression trees, clustering algorithms, neural nets, and Bayesian nets.

  and retrieving objects managed by the agora and for the transactions by the agora's patrons and customers, they are not well suited for extracting the patterns, associations, and anomalies listed above. This is the role of data mining software now being developed. Data mining software has by and large been developed for mining small amounts of data that has been collected and centralized. It is important to develop data mining algorithms which are scalable and which work with the distributed nature of the agora.

4.3.3 Audit and Control

Electronic transactions will be recorded to enable audit and control. Developing tools for automatic analysis of the audit trail is an important technical challenge in machine learning and data mining. The objectives of clients and servers in analyzing the audit trail are to increase revenues and reduce costs. For example, the tools should enable the discovery of marketing opportunities, and the discovery of pricing and negotiation weaknesses. Furthermore, the tools may enable automatic modification of these strategies to incorporate the lessons learned. Otherwise, there should be a seamless feed of these lessons into decision support systems and problem solving tools.

4.3.4 Ontolgies

In EC/DL systems it is desirable that users be able to query across domains. In addition, a user familiar with one domain should be able to leverage that knowledge for use in other domains. Ontology is defined as the set of terms and relationships used in a domain, denoting concepts and objects, often ambiguous among domains [10].

Presently, ontologies are being constructed in a number of domains for a wide variety of purposes. Some, such as the I3 glossary itself, are used as a domain specific dictionary/thesaurus while others are used as a basis for exchanging information between information systems designed around differing domains [17].

Ontologies are described in some type of representation language. Some examples of this include LOOM, Epikit, Algernon and KIF. As several researchers have proposed such languages, a means to translate (map) between them is desirable. In [10], Gruber describes a portable ontology specification called Ontolingua that can be used as a translation tool between ontologies described in several representation languages. With such tools, the mechanics of mapping between ontologies is made easier.

¿From a conceptual standpoint, finding the appropriate mappings between ontologies can make use of at least two possible scenarios:

- Specialists familiar with multiple domains can point out the relationships between ontologies. This is highly labor intensive and may take longer to achieve that the initial creation of an ontology. For example, given n ontologies, a new term introduced into one ontology would require specification of $n-1$ new relationships.

- We can also envision a semi-automated mapping tool that is capable of deducing relationships between ontologies based on relationships within ontologies. In Figure ??, it should be sufficient to specify `GIS.layer IS-A Medicine.slice`. What would naturally follow is `GIS.band IS-A Medicine.slice`. The semantics for other types of relationships may be exploited in a similar fashion.

Given a means to map ontologies from multiple domains, the question remains as to the utility of this "new" ontology. Measuring the quality of any ontology is subjective to the creators of that ontology. What is of particular interest is what may be "lost" when mapping between ontologies or merging ontologies. In [10], Gruber points out that Ontolingua is indeed a "lowest common denominator" approach that may not be able to translate complex expressions from all representation languages. When mapping between ontologies, the representation used must be robust enough to accommodate the representations of the individual ontologies.

4.4 Using information

4.4.1 User Interface

The user interface is an important component of an EC/DL system. It must incorporate a wide variety of techniques to afford rich interaction between users and the information they seek. In EC/DL environment a large amount of data spread through a number of resources necessitates intuitive interfaces for users to query and retrieve information. The ability to smoothly change the user's perspective from high-level summarization information down to a specific paragraph of a document or scene from a film remains a challenge to user interface researchers.

Users of a traditional library usually want to read a book or watch a videotape; other uses are rare. With digital technology, it now becomes possible to listen to a book being read, watch a video of a musical performance alongside the original score, or hold a mechanical hand as it forms American Sign Language. Other possible uses are highly personaf individuals may dream up many distinct variations. An EC/DL's presentation systems must be flexible and highly customizable. They must also be aware of the output hardware's capabilities and limitations, automatically adjusting to deliver the best possible presentation quality at all times.

An EC/DL environment should provide authors and information providers with powerful tools that take into consideration the following:

- Author perspectives - the author specifies different manifestations based on his/her conceptual constraints. For example, the author of a medical image object may pose a constraint that this object can only be played on high resolution video devices because a low resolution device won't be able to render the detail.

- Functional perspectives - Hardware and software environments preclude certain manifestations. For example, an information appliance can not render a video clip without having a video card. Another example is that a user may have a high resolution video device such as

SVGA; however, the object can not be played if there is no a proper software such as MPEG or JPEG.

- User profile and preference perspectives - that define a manifestation based on users' profiles and preferences.

4.4.2 Financial instruments

Examples of financial instruments include bidding, purchase orders, digital cash, extension of credit. An important research issue is the development of a technology for electronic transactions, such as trade, with complex objectives. A typical scenario for electronic trade over Internet involves multiple suppliers offering products and services and large number of bidding consumers. An example of a complex objective for, say, a chemicals manufacturer having constraints on resources, is: find materials' suppliers and a set of electronic bids that can be produced under the constraints and would maximize the profit.

The most natural starting point is Constraint Databases that have emerged to incorporate data captured by constraints in databases and are motivated by the fact that constraints are an ideal media for complex modeling and a data-type which is (1) uniform, (2) expressive, (3) compact, and (4) deeply optimizable for many specific domains.

One important challenge is the development of flexible operations paradigm for electronic traders and analysts. The data model ought to be flexible and to enable update the database and the schema dynamically without involvement of a "centralized" system administrator. Note, different classifications of products are part of the schema, but are definitely something that would be often updated. The querying paradigm should be suitable for a (possibly sophisticated) analyst, which is capable of understanding and describing complex trade objectives, but is not a programmer. Not only ad-hock queries should be allowed, but rather filtering queries that will be periodically applied with new information available from the Internet.

A probably more challenging is the question of query evaluation, which include indexing and filtering, constraint algebra algorithms, and global optimization. The practical feasibility of electronic trade systems rely heavily of the ability to deeply optimize queries. The algorithms for constraint algebra are significantly different from those in relational algebra, although many ideas are still applicable. The traditional database optimization, both compile-time algebraic simplification and run-time cost-based approach are not adequate in constraint databases dealing with information over Internet. The main reason is that both constraint manipulation and access to data over Internet is computationally expensive; thus a fine balance should be found between the cost of each. Even cost estimation of evaluation plans may be infeasible without spending considerable amount of work on constraint manipulation. For filtering queries, dynamic views maintenance in presence of large constraint data sets presents a particular difficulty.

Implementation of a prototype electronic trade system and performing a case study would be of great interest. This can be done on top of an object-oriented constraint database system such as C^3 which has been developed at George Mason University.

The above is simply an example of a particular approach, Many other approaches are possible. Whichever ones ends being used, it appears that electronic payment mechanisms are essential for such methods to be implemented. With such mechanisms in place, we expect to see Internet as "meta-computer" on which computing, storage and communications resource will be freely traded on this new "electronic market-place". This will finally accomplish the goals of information revolution: defeating the tyranny of geography.

4.5 Securing information

4.5.1 Security

As the number and availability of digital libraries grow, information security becomes both more critical and more challenging – more critical because of the wide range of information and the numbers of ad hoc users, more challenging because the characteristics of digital libraries differ from traditional information systems and call for new types of controls. The complexity of specifying EC/DL security policies, the difficulty of implementing them over existing systems, and, most importantly, the need to cope with differences and interactions among autonomous providers require innovative solutions.

Digital libraries are generally distributed, but they differ from other distributed systems in that they often consist of changing collections of documents provided by a changing group of autonomous servers. While information providers may be concerned about retaining control over their documents, the same providers are also less likely to agree to abide by a single, uncustomizable set of externally-imposed policies. Traditional multilevel information security techniques also assume a known set of users, each cleared to specific levels and subject to the same policies and laws. Since these assumptions do not apply to users of digital libraries, information providers may wish to attach terms and conditions (including copyright controls, fees for use, modification restrictions, etc.) to documents when they are made available [15, 4] Furthermore, digital libraries are often built over legacy systems, possibly with specialized front ends for retrieval, but providing little in the way of sophisticated underlying authorization mechanisms. Thus, even some of the capabilities now provided in commercial secure database management systems must be retrofitted.

The greatest design challenges for secure digital library systems lie in the formulation, specification, and enforcement of comprehensive data protection policies. These policies must take into account the differing concerns of authors, publishers, librarians, and users, and they must offer three varieties of security: confidentiality, authenticity, and integrity.

Confidentiality is perhaps the most basic security concern in all systems involving sensitive or valuable data, and the problem is exacerbated in digital library environments.

- Information providers will not be able to vet all users ahead of time.

- Providers may need to rely on other organizations' judgments that specific users are authorized to see particular information.

- Confidentiality for EC/DL systems necessarily includes not just release, but terms of release such as billing. The release policy for a document may not just determine whether a document may be released, but what charge should be assessed and who should get the bill.

- Copies of an object must generally remain subject to the same release policies as the original, while excerpts may or may not involve different policies. Note that the determination of which release policy is relevant may also depend on the user and the nature of the task. For example, very short excerpts or excerpts by reviewers or educational institutions may not be subject to the same fees or restrictions as other retrievals.

- In EC/DL environment, time information may be a crucial part of confidentiality. A price list or commodity report may be highly confidential until a certain time and completely available after that time. For some new research in this area, see [2].

- Just as digital library researchers are actively engaged in improving the indexing and retrieval of text, image, video, and audio data, research is needed to develop feasible techniques that determine and describe the sensitivity of such data objects. The operation used over such data may also be important; for example, sensitive image data may be made more widely available at lower resolutions.

- Information sources or author information may be more sensitive than the documents themselves (e.g., press leaks and background information).

- Users are tremendously concerned about the confidentiality of usage patterns, an especially difficult problem when itemized bills are needed. Authenticity can be ensured to a large degree by the technologies discussed earlier. However, for digital libraries, the problem shifts to specifying and enforcing policies that reflect the types and uses of digital library data.

- If the source of a document is more sensitive than the document itself, a person granted access to the document may not be able to verify authenticity (perhaps even more crucial precisely in the cases when the source is extremely sensitive).

- Time plays a part in authenticity as well as confidentiality. It is often essential to know, for instance, that the version of a document being read was actually completed on a certain date (for example, an authentic insurance policy must have been written before a claim is made). As digital library authenticity policies are specified, other types of information pertaining to authenticity may also be identified. Confidentiality and performance issues, however, will affect the final choice.

- Users and information providers may be interested in a "history" of authenticity – verifying who was responsible for revisions, summaries, excerpts, paraphrases, citations, etc. of previous documents.

- Copies may be made in various ways; certain operations may be certified to preserve authenticity and authenticity information, such as checksums over the author and the document, while others may not.

- The authenticity of a composite document must include information about the authenticity of the components and of the links. Authenticity of links depends not only on the existence of the link, but on the type of link (e.g., whether a name refers to the author, the provider, or the reviewer of the document) and the date.

  Integrity refers to protecting documents (and ancillary information such as the source) from unauthorized modification. Integrity is similar to authenticity, in that a document's authenticity depends in part on preserving its integrity. Therefore, issues similar to those raised for authenticity apply to integrity, but digital library usage also raises some additional integrity problems.

- Users may wish to know not only that all previous changes to a document were authorized, but also what those changes were (i.e., which operations or tasks were performed). In some situations, however, the sources, dates, or types of changes may be sensitive (e.g., changing budgets, ghost writing, etc.), so integrity/authentication must be balanced against confidentiality.

- To preserve integrity, derived data must inherit data protection policies – unauthorized modification of an excerpt can cause just as much trouble as unauthorized modification of the original document.

The secure EC/DL system, then, must address the concerns of its constituents with respect to confidentiality, authenticity, and integrity. Given the volume and complexity of data, the system design must balance security concerns against responsiveness and performance. While these challenges are great, perhaps the thorniest issue is the need to deal with policy interactions across distributed providers. Providers may have different concerns and motivations, and they may use variants of access operations. Providers' security policies certainly will differ. Conflicts among policies must be handled in reasonable ways, without unduly denying or compromising service. Security policies and enforcement mechanisms must be explicitly considered as one aspect of heterogeneity to be managed in digital libraries.

Several existing research areas and emerging technologies appear promising for at least partially addressing the security needs of EC/DL, though they will need adaptation or extension for the unique characteristics of this environment. Digital signatures, checksums, and encryption (see, for example, [16] or [1] will provide support for some critical aspects of authenticity and confidentiality. Security features that will be refined as part of the maturation of object database technology will help with protecting composite and multimedia objects in terms of extensible operation sets. Work being done in security for federated database systems will help with issues of the interaction of autonomous policies (see, for example, [4, 7].

Recent advancement in holographic storage and retrieving technology offers a new and promising security verification system for identification and/or credit cards. By utilizing a pseudo-randomly generated phase mask which stores, in a nearly invisible fashion, biometrics information and secret code on ID cards, an all-optical correlation can quickly and reliably separate authentic IDs from counterfeits.

4.5.2 Intellectual property rights

Commercial transactions involving information units (e.g. books, articles, movies) will necessitate the transfer and enforcement some/all intellectual property rights. For example, the sale of a book may or may not allow the buyer to reproduce parts of the book for profit. We will need to develop schemes for the enforcement of intellectual property rights, and the detection of violations. Some progress on this front is already being made. For example, several recent works propose methods of efficient automatic detection of "suspicious" similarity between documents.

4.6 Universal access

4.6.1 Interoperability

An EC/DL system should interoperate with legacy and other systems such as database management systems and workflow systems. For example, EC/DL clients and servers will need to retrieve from and update information in corporate databases. Workflow systems will be responsible for monitor and control of work assignments in organizations. While flowing through an organization a work-assignment may involve accessing the global digital library, or buying/selling a product. Thus EC/DL clients and servers may need to trigger and be triggered by workflow systems; data will also have to be exchanged between the two types of systems. Mediation and heterogeneous databases are two promising research directions currently being explored to address interoperability.

The complex interaction among the heterogeneous systems will need to be managed by facilities for schema integration and standard protocols.

These facilities will provide users and programmers with some form of a global schema of available information, albeit in a very restricted domain. Standard protocols will enable a client to search for the best provider of a service or product, and will enable the provider to maximize its profit. Http is an example of such a protocol, but higher layers will have to be built to give the infrastructure more intelligence, and elevate the abstraction it provides. International commerce will greatly be facilitated by some form of automatic translation between languages, at least in the restricted domain of electronic commerce.

Little attention has been paid to the content of libraries that do not lend themselves to digitization. Items such as a plaque or trophy, player piano scrolls and other so called *legacy* artifacts are often within the domain of libraries. In this sense, the line between museums and libraries is often difficult to draw concisely [18].

Certainly it is possible to videotape such artifacts from a number of angles and store this within the digital library. However, for many items, this would not afford the interaction necessary for some scholarly research. In the foreseeable future, it will therefore be necessary to incorporate these artifacts within the scope of digital libraries.

4.6.2 *Format Conversion and Data Interchange*

. Conversion services to promote interoperation is an essential service. Examples of such a service include support interoperation among different Electronic Data Interchange (EDI) standards and other electronic forms.

4.6.3 Cost Management

Today, information-service providers and consumers use cost models that are time based, or request based, or a combination of the two. In a time-based model the customer pays for unlimited access to the service for a time-unit (say a month); cable television uses this model. In a request based model the customer pays per request (e.g. news-stand purchase of a newspaper). Once a cost model is selected by an information provider, it remains fixed until it is changed manually. Given limited resources and the marketing opportunities presented to information providers by the new electronic medium, this solution is unsatisfactory. Similarly, information consumers have unprecedented capabilities for comparison shopping that results in cost minimization. Therefore, we need to develop algorithms and systems that will enable cost minimization for information consumers, and exploitation of new marketing opportunities for information providers.

For information providers, we need to develop algorithms and systems for automatically determining the price of information. These systems should dynamically adapt the prices, and maybe even the cost models, to supply and demand. What are the parameters and primitive functions of an information-pricing system? A related issue is how can some or all intellectual property rights be transferred automatically from the owner to subcontractors and consumers? When to do so, and how to charge for these rights? How can intellectual property rights be enforced?

Another issue is to study cost models from the consumers' point of view. Consumers will receive bills from multiple information providers. In many cases these charges can be optimized, and how to do so depends on the available providers of a piece of information, their cost models, and the access patterns and needs. Therefore, we need to develop systems that will assist the consumer in cost management. For example, consider a consumer reading the New York Times. She can subscribe to receive the newspaper every day, and in this case the price per copy is lower than the (electronic) news-stand price. However, the the consumer may pay for issues that she does not have the time to read. On the other hand, the news-stand price may be higher, but the consumer buys the access only when she is certain to read the paper. Which is the better access protocol? Obviously, the answer can be computed based on the subscription versus news-stand price, and based on the probability that the consumer reads the paper on a given day; and this probability can be computed based on past experience, i.e., a cost optimization system can "learn" the access pattern of the user.

Therefore, from the consumer's point of view, the optimal provider and/or access protocol depends on the access pattern, and on the various cost models and providers of a particular piece of information. In [21, 19] we analyzed several algorithms that select a provider, cost model, and access protocol for cost optimization. Another factor that determines the consumer cost is the needed timeliness of the information (e.g., if the consumer can tolerate data that is 10 minutes out-of-date, then both, the communication cost and the access cost may be reduced). In [?] we analyzed some aspects of this problem.

Finally, for both, consumers and information providers it is important to develop algorithms and systems that retrospectively examine pricing and purchase decisions; they should be able make or recommend strategy changes.

4.6.4 Resource management

In a multi-user environment, e.g., time-shared systems, Internet, etc. resources are traditionally managed based on concepts of "fairness", and rigid "priority" structure, As the demand exceeds supply, performance degrades uniformly. In economic terms, this is the essence of "communist" economy. Such economic system is obviously the best if supply exceeds demand. Otherwise, competitive open-market environment performs better.

We are currently living in the "communism" era, where payments for electronic services is very uncommon. However, we can observe that demand for resources (e.g., number of users on the Internet) grows exponentially while supply of resources (e.g., total bandwidth available) exhibits a much slower growth. This calls for revision of our approach to research allocation on the Internet (or Global Information Infrastructure). More specifically, we call for the "information-perestroyka", namely transition from rigid central planning to open markets of electronic resources (computation, space, communication) and services (software, information). We argue that this is the next step is crucial to reap the fruit after of the "information-revolution" for further information see http://www.cs.jhu.edu/ baruch/META-COMPUTING/index.html.

4.6.5 Networking

Increased demands for networking bandwidth come from two main fronts. First, the number of digital library users will undoubtedly increase. If the Internet is any indication, exponential growth in the number of users will be the rule. Second, as the delivery of multimedia data becomes the norm, the demands for high bandwidth increase. However, high bandwidth, in and of itself, is not enough to support EC/DL systems. The intelligent use of bandwidth and the ability to guarantee bandwidth for a given time period are also required. Today's open networking standards such as TCP/IP and the ever-growing Internet make it clear that successful digital libraries must be built on an open, interoperable networking infrastructure. Current CE/DL systems may be run exclusively on a single computer, on several computers connected on a LAN, or on a large number of computers spread out over a wide area network. Delivery systems that require high bandwidth such as video and image libraries are predominantly installed using LANs that run at 10 to 100 Mbits per second. In contrast, the Internet's major backbones run at 1.5 Mbps to 150 Mbps, while links to individual organizations fall in the 56 Kbps to 1.5 Mbps range. Individual users typically connect to the

Internet through service providers, local universities, or other organizations at 2.4 Kbps to 28.8 Kbps.

To make the next generation of EC/DL systems widely accessible, the following technologies will play a major role:

- Wireless network. The wireless technology allows rapid and economical deployment of EC/DL systems to both developed and rural areas, to households, offices and mobile stations. By combining with satellite (free space) and advanced fiber optical networks (in ground), the wireless network can be potentially assembled as a inter-city, inter-country or even global network.

- Advanced optical network. One of the key communication system resources for accessibility is bandwidth. The emerging ultra high bandwidth fiber optical communication technologies (all-optical time domain multiplexing, TDM, and wavelength division multiplexing, WDM,) will greatly expand the communication bandwidth (by several orders of magnitude). The all-optical switching potential inherited in the WDM network systems will offer vast capacity and efficient connectivity to provide bandwidth on demand.

- The ultra-high bandwidth optical network systems, implemented with new communication protocols such as ATM, are essential in order to meet the future requirements of high information transmission volume with guaranteed quality of service. The WDM optical network will parallelize the network switching and routing functions, thus greatly simplify the network management tasks. For a private network, a WDM system can also be operated in a dark fiber with protocol transparency.

- Fiber to home. The high bandwidth communication service telephone and/or cable companies plan to offer to their customers will greatly enhance the connectivity of most household. The quality of service of EC/DL systems will benefit greatly from this increase of connectivity. Additional services can also be realized with this infrastructure enhancement.

- Communication system: The advanced optical communication network being researched on will offer large capacity, redundancy, and reconfigurability. The WDM systems are distributed networking systems which offer more flexibility in re-routing and reconfiguration and thus are more reliable as well.

4.7 Socio-economic Issues

Six major issue areas are identified which directly flow from the impact of information technology on the world's economy, society and political communities. Our goal in raising these for continued discussion is to insure the beneficial implementation of EC/DL systems. We are not content to see our technology revolutionize the way the world works without also expressing our ideas and recommendations on ways to keep technology humane.

Many of the social and economic issues identified can be addressed by new and modified technology. We can create technical answers to many of the public policy problems outlined in the following comments. We want to know from our colleagues in other disciplines and our fellow citizens what these problems are. We seek their cooperation in finding solutions.

4.7.1 Economics:

The world of trade and business is changing dramatically through electronic commerce. Electronic commerce, or cyber business, is much more than simply doing electronically what already goes on in ordinary business. Information technology applied to this global business environment creates new jobs, and eliminates others; it creates new concepts of business, and challenges old ways to buying and selling; it calls for new ways to pay for goods and services, and threatens monetary regulatory schemes.

Knowledge and information, now stored in inaccessible libraries, takes on new economic value through cyber distribution networks. This raises serious policy questions about the costs of accessing information, and about the intellectual property rights of authors and creators of new forms of information. Who should have access to knowledge on-line? Who should pay?

Technology is reshaping the patterns of work. 30 million Americans already spend more than 30 hours a week producing real work from their home offices. Different patterns of economic life, created by computers and networks, is impacting the very structures of communities. Most traditional downtown office buildings are nothing more than multi-story filing cabinets. Technology will force major shifts in employment and the design of jobs. New skills and training are necessary right now.

International trade will be changed as increasing numbers of business transactions are handled through electronic networks instead of through mail and personal travel. Nations will be frustrated in their need to collect customs or regulate what comes into their economies.

New systems for payment will be required. The cashless economy, without the flow of traditional national currencies, will be a major challenge to regulators of the world's monetary systems. Can the networks and systems handle the dramatic increase of retail transactions? How can we preserve the psychological value which people in the marketplace attribute to hard money?

Insuring that all people everywhere have access to the global information infrastructure is a major public policy issue. Currently, universal POTS (plain old telephone service) is forced through government economic regulatory requirements. But in a world where nearly two-thirds of the people have never even made their first phone call (source: International Telecommunications Union-ITU), how do we get to global access to the knowledge which we believe will help them overcome ignorance and poverty?

One interesting question arising in connection with the appearance of network-based information/purchasing services is who should pay for the resulting abundance of readily available information. Buyers or s sellers? Traditionally, vendors have paid to reach and service their customers without (explicitly) charging customers for any part of the information transmitted. Traditional brokers normally inflate the cost of the product sold to recover cost rather than account for the acquisition cost as a separate line item on the customer's invoice. access to information databases like easy Sabre and Apollo are sometimes accessed via a connect-time charge for the customer (Eaasy Sabre on Prodigy) and sometimes not (Apollo on the Internet Travel Network). This is clearly a new business arena in search of the most appropriate business model. Since most information technology requires electricity to operate, it is imperative that we concern ourselves with energy resource questions, the energy efficiency of technical tools.

4.7.2 Human Resources

Since information technology usage presumes literacy, not only in computer operation, but in basic language,Information technology applied wisely will give us a new opportunity to address the global scandal of 40 to 60 percent absolute illiteracy in too many countries, and the 20% functional illiteracy in industrialized 0% countries. (World Bank source). We believe that information technology properly deployed in educational services can solve this problem.

The technology itself must be user friendly. Our challenge is to design tools of access to this global knowledge network for everyone, including those with intellectual and physical disabilities.

Workers will be challenged by a shift in the nature of work itself. Training for new skills will be needed for new jobs. "Just in time knowledge," and "lifelong learning" will be facilitated by information technology. The impact on the workplace is already apparent. In the U.S., almost 10% of economic activity is now generated in the information sector. Out of all U.S. workers, 60% have jobs which depend on the information they generate and receive on advanced information networks (ie. electronic commerce services). In terms of the future, it is estimated that 8 out of 10 new jobs created are in the information-intensive sectors of the U.S. economy (statistics from Electronic Commerce Innovation Centre; article "An Introduction to Electronic Commerce)

Global use of information technology will allow for global cultural awareness and communications. It will also facilitate preservation of small community identity. It is essential to explore the impact of technology on human inter-relationships. Some believe that computers and networks will isolate people; other see them as tools for reaching out and touching even more people.

Because the filtering process of traditional editing and publishing of creative works is absent from networks people are faced with the unwanted assault of undesirable information and images. The issue of the quality of information available must be addressed. The protection of one's personal privacy, and one's family and community values is more than just a legal right, it is essential to psychological security and stability. We need to understand this basic human element and build technology to serve and achieve the highest goals.

4.7.3 Legal and Government

There is not universal agreement in the laws of all nations on the ideas of freedom of speech and the press, nor on the protection of privacy of information about individuals. Information, ideas and images placed on the networks in the United States, where these are constitutionally protected, may be considered offensive or illegal in another nation with different laws. Confidential information placed on the internet can be accessed by others around the world. This creates questions concerning legal protections and governance of the Internet. Technology can provide tools for such protections within the context of varying legal systems. We should find a balance between unwanted government censorship of the Internet, and cyber anarchy which violates individual privacy and community values. The very technology upon which all networks depend operates within exacting standards, and absolute rules of science. Certainly, we can arrive at rules of governance and internethics for the Internet which are balanced and civil.

Governments must negotiate the policy framework within which electronic commerce can take place across international boundaries. Questions of fraud, business crimes, tax collections and currency management must be resolved both through policy decisions and application of technical tools.

Governments have to face the issue of access to information networks for citizens of their country. They must resolve questions of what content should be available, and who will pay for it.

Richard Power's arcticle "CSI Special Report on Electronic Commerce: Treasure of the Sierra Madre?" in the Spring 1996 edition of Computer Security Journal, presents some good references to challenges governments face as to electronic commerce. He cites an interview with former federal prosecutor, Andrew Grosso, who currently is a private computer security lawyer in Washington D.C., on this issue. "Sovereignty is based on the control of borders and currency. But the Internet challenges both of these. You can't regulate people beyond your borders. And, in fact, with the availability of anonymous servers located overseas, it has become very difficult to regulate those within your own borders. Likewise, the concept of digital cash stresses government control over the flow of currency." On the economic side of these legal issues, Grosso sees a potential loss of revenue for state and local governments. "If a consumer purchases a large ticket item like a car or refrigerator using digital cash through an account with an out-of-state provider or an anonymous server overseas, how will the vendor know whether or not to pay sales tax?" A coordinated international agreement must be reached on issues like taxation and regulation. These are some issues that need to be addressed if the growing electronic community is to reach critical mass.

Grosso also mentions potential problems for law enforcement and an opportunity for organized crime. "Digital cash will make money laundering a lot easier. Currently, criminal organizations require financial institutions to launder large amounts of cash. You just can't move it physically. Couriers running arround with millions of dollars in their luggage are noticed. They can't make it past the airport x-ray machines. However, a single electronic commerce server could easily pump out a $1 dollar transaction every millisecond. That's $1,000 a second and over $300,000 an hour. Faster servers can pump out millions of dollars an hour. How could law enforcement agencies monitor these transactions? They would have to turn the Internet into one giant surveillance system. They would have to monitor every single financial transaction, which means they would have to monitor every single digital message to see if its a financial transaction." These are some of the legal challenges governments face in this area.

4.7.4 Society

The technical community must fast questions about the ergonomics of information technology. Absolute consumer safety and universal consumer usage must be assured through design of the technical tools.

Cultural and multilingual issues arise with the ubiquitous nature of the Internet. Several nations are resisting allowing access to the Internet because of the dominance of English on the net. Technology can provide an answer to allow access to those with other language skills and heritage. We must find a way to both benefit from the global reach and connection of the Internet, and personalize its services to varieties of people with different languages and cultural backgrounds..

4.7.5 Privacy and Security

"By almost all accounts, security is the leading barrier to widespread commerce on the Internet." (Anish Bhimani, "Securing the Commercial Internet," Communications of the ACM) Global networks will require international agreements to protect the privacy and security of individuals and of commercial transactions. However, privacy and security will only be as strong as the technology developed to provide it. We must strive to increase these protections. Security systems on the Internet have become more than a technological issue, though.

Current export regulations on encryption systems presents a challenge for both policy makers and technologists. Existing regulations significantly inhibit the deployment of a safe, worldwide network infrastructure for electronic commerce and digital libraries. We must decide if this is just a temporary legal speed bump that will go away with time and technology, or does a serious lobbying effort need to be undertaken by industry and technologists together so that up to the minute security technologies can be made available globally. The CommerceNet group published an article “The Need for a Revised U.S. Cryptographic Policy” which provides evidence that current policies are harmful to companies marketing encryption products. (ie. triple DES is the ISO standard for encryption in certain banking funds transfer applications. U.S. export regulations make U.S. banks unable to implement the international standard for overseas transactions.) The article asserts that strong, reliable and robust encryption applications exist today to provide security service in the marketplace. There are existing forces preventing their wide scale deployment. The government argument is that high tech, easily accessible cryptography is a threat to national security and crime prevention. The concern is “that encryption will be used to hide illegal activities, prevent authorized tapping of communications lines, and otherwise mask terrorist acts.” It appears that Europe is following the United States' lead in officially prohibiting encryption systems beyond the control of governments. The European Council has passed a similar initiative to the current U.S. policy for these systems. (EDI World Institute, Nov., 1995) This is definitely one area where technologists and policy makers need to work together to reach a reasonable balance between government agencies and the market need of powerful cryptography for electronic commerce.

4.7.6 Multi Discipline and Interaction

The issues raised and the problems to be solved in building the global information infrastructure are a mix of technology, public policy and culture. Finding solutions will require the cooperation of many disciplines, i.e. law, economics, business, education, sociology, psychology, entertainment, and computer science, etc. Accomplishing the information revolution need not be accidental. We can cooperate to develop solutions before avoidable disasters occur.

We as the computer society must open dialogue with others in the variety of disciplines. We must teach others about technology, and learn ourselves about other domains of knowledge. We must make ourselves aware of public policy developments, and seek to influence these decisions. We should offer spokespersons to conferences on relevant issues.

Electronic commerce and digital libraries represent two related and parallel techno-developments which will drive decision making in all the broader issues listed above. The desire to do business and the search for knowledge are two fundamental human activities and values which cut across all cultures. We are prepared to engage in the dialogue to build the global information society.

5 Conclusions

To be completed later

References

[1] N. R. Adam, B. S. Fordham, and Y. Yesha. Some Key Issues in Database Systems in a Digital Library Setting. In *Lecture Notes in Computer Science: Digital Libraries Current Issues*, volume 916, pages 9 – 20. Springer Verlag, May 1994.

[2] N. R. Adam, M. Halem, and S. Naqvi. Promising Research Direction in Digital Libraries. In *Lecture Notes in Computer Science: Digital Libraries Current Issues*, volume 916, pages 21 – 30. Springer Verlag, May 1994.

[3] H. Berghel and L. O'Gorman. Protecting ownership rights through digital watermarking. *IEEE Computer*, 29(7):101 – 103, July 1996.

[4] E. Bertino, C. Bettini, E. Ferrari, and P. Samarati. A Temporal Access Control Mechanism for Database Systems. *Transaction on Knowledge and Data Engineering*, 8(1):67 – 80, February 1996.

[5] Biederman. Recognition-by-components: a theory of human image understanding. *Psychological Review*, 94(2):115 – 147, 1987.

[6] B. T. Blaustein, A. McCollum, C. D.and Rosenthal, K. P. Smith, and L. Notargiacomo. Autonomy and Confidentiality: Secure Federated Data Management, June 1995.

[7] H. Wactlar et al. Intelligent Access to Digital Video: Informedia Project. *IEEE Computer*, pages 46 – 52, September 1995.

[8] W. Niblack et al. The QBIC Project: Querying Images by Content Using Color, Texture, and Shape, 1993.

[9] W. Niblack et al. Integrity Constraints in Federated Databases, 1996.

[10] et al. Faloutsos. Efficient and effective querying by image content. *Journal of Intelligent Information Systems*, pages 231 – 262, 1994.

[11] T. R. Gruber. A translation approach to portable ontology specifications. *Knowledge Acquisition*, 2(5):199 – 220, 1993.

[12] R. M. Haralick and L. G. Shapiro. *Computer and Robot Vision, Volume I and II.* Addison-Wesley, 1993.

[13] David J. Heeger and Patrick C. Teo. A Model of Perceptual Image Fidelity, October 1995.

[14] Y. Huang, P. Sistla, and O. Wolfson. Data Replication for Mobile Computers, May 1994.

[15] H. V. Jagadish. A retrieval technique for similar shapes, May 1991.

[16] Jain, Murthy, and Chen. Similarity Measures for Image Databases, February 1995.

[17] G. Jung and V. Gudivada. Adaptive Query Reformulation in Attribute-Based Image Retrieval, 1994.

[18] Robert
Kahn and Robert Wilensky. A Framework for Distributed Digital Object Services. *http://WWW.CNRI.Reston.VA.US/home/cstr/arch/k-w.html*, May 1995.

[19] Charlie Kaufman, Radia Perlman, and Mike Speciner. *Network Security.* Prentice Hall, 1995.

[20] F. Lehmann. Machine-Netgotiated Ontology-Based EDI, November 1994.

[21] Michael Lesk. Which Way to the Future? The Control of Scholarly Publication. In *Lecture Notes in Computer Science: Digital Libraries Current Issues*, volume 916, pages 33 – 50. Springer Verlag, May 1994.

[22] D. M. Levy and C. C. Marshal. Going Digital: A Look at Assumptions Underlying Digital Libraries. *Communications of the ACM*, 38(4):77 – 84, April 1995.

[23] V. E. Ogle and M. Stonebraker. Chabot: Retrieval from a Relational Database of Images. *IEEE Computer*, pages 40 – 48, September 1995.

[24] P. Sistla, O. Wolfson, S. Dao, K. Narayanan, and R. Raj. An Architecture for Consumer-Oriented Online Database Services, February 1996.