Communications of the Association for Information Systems

Volume 4 Paper

July 2000

Web Based Knowledge Infrastructures for the Sciences: An Adaptive Document

Alexander Hars
University of Southern California, alexander.hars@marshall.usc.edu

Recommended Citation
Hars, A. (2000). Web Based Knowledge Infrastructures for the Sciences: An Adaptive Document. Communications of the Association for Information Systems, 4, pp-pp. https://doi.org/10.17705/1CAIS.00401

Communications of the Association for Information Systems

Volume 4, Article 1
July 2000

WEB BASED KNOWLEDGE INFRASTRUCTURES FOR THE SCIENCES: AN ADAPTIVE DOCUMENT

Alexander Hars
Marshall School of Business
University of Southern California
Hars@usc.edu

KNOWLEDGE MANAGEMENT

WEB BASED KNOWLEDGE INFRASTRUCTURES FOR THE SCIENCES: AN ADAPTIVE DOCUMENT

Alexander Hars
Marshall School of Business
University of Southern California
hars@usc.edu

ABSTRACT

Scientific knowledge is increasingly being stored online. A large number of infrastructures that provide access to scientific knowledge are now available on the Internet. They range from online journals to collaboratories and logic servers. This article examines a variety of such infrastructures and derives implications for their further evolution. It assesses their impact on the way that scientists will create, organize and integrate knowledge. In parallel, the article illustrates how online content may become more adaptive and structured. The text consists of individually marked sections that are assembled dynamically to the needs of each reader on different levels of detail.

Keywords: Knowledge management, knowledge infrastructure, online journal, collaboratory, impacts of information technology, adaptive document, XML

Author's note: The adaptive features of this article are only available in the HTML document. This PDF version contains the superset of all versions available in HTML. It is very detailed and may only appeal to the reader with a strong interest in all aspects of the topic. The *extended abstract* or *main article* versions in HTML are more appropriate for the casual reader.

Legend: Red=Instructions; Blue=Theory/Method; Green = Application; Dark Red=Further Detail

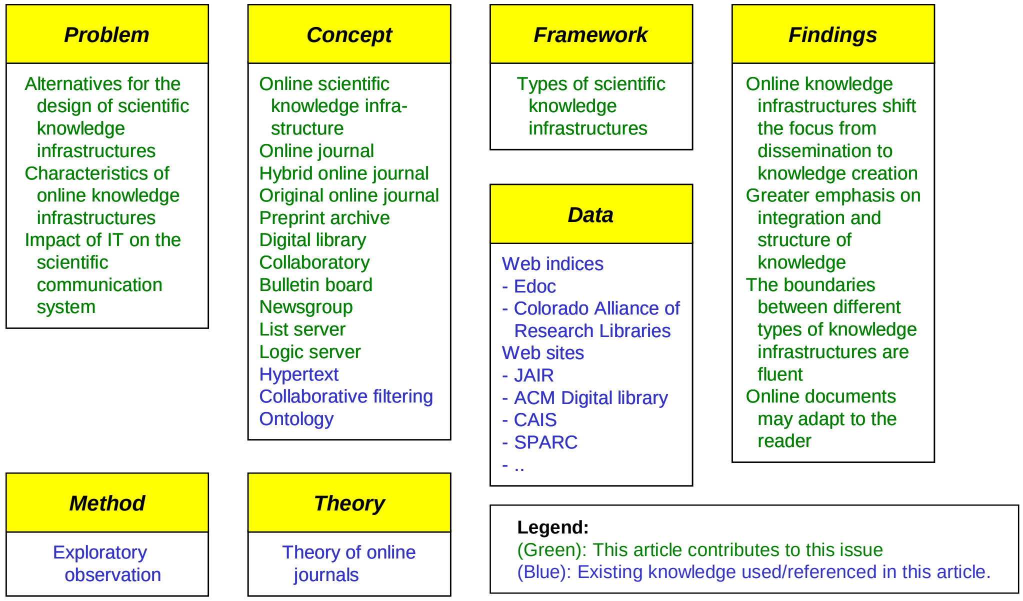


Figure 1: Table of Contents (by categories of scientific knowledge)

I. INTRODUCTION

Scientific knowledge is increasingly being stored online. A large number of publishers, scientific associations and individual researchers are currently launching online sites that are dedicated to scientific knowledge. This leads to the following three problems:

ALTERNATIVES FOR THE DESIGN OF ONLINE SCIENTIFIC KNOWLEDGE INFRASTRUCTURES.

The designers of internet-based knowledge infrastructures spend considerable effort in experimenting with novel approaches. They need to understand the range of approaches that have been implemented successfully by other pioneers. Thus it is necessary to describe the variety of ideas and conceptualizations on which current knowledge infrastructures are based. This description enables designers to make more informed decisions about the structure used to represent scientific knowledge and about online support for the processes of creating, disseminating, integrating and deleting knowledge.

CHARACTERISTICS OF ONLINE INFRASTRUCTURES.

The identification of design alternatives requires an analysis of the categories and characteristics of existing online scientific knowledge infrastructures. The literature currently distinguishes between several categories of infrastructures such as online journals, collaboratories, digital libraries etc. Exemplars for each of these categories can be found on the Internet; their key characteristics need to be aggregated and compared to the characteristics of other categories. An alternative to such an empirical approach is a deductive approach that systematically derives insights about online knowledge infrastructures by examining related concepts. Theoretical accounts of the traditional publishing system could be used to examine the impact of information technology on each element of the traditional system. The individual changes can then be aggregated into a consistent picture of an online publishing system that can be used as one approximation for the emerging online infrastructures. The evolution of online infrastructures could then be compared against the theoretical conceptualization. An example of this approach is presented in Hars (1999). The advantage of the deductive approach is that it is not limited by the immaturity and rapid evolution of current information technology. On the other hand it requires assumptions about the adoption of information technology that may later turn out to be unrealistic.

IMPACT OF IT ON THE SCIENTIFIC COMMUNICATION SYSTEM

The current evolution towards online infrastructures leads to the more general problem of forecasting the impact which information technology may have on the scientific communication system. Information technology may lead to fundamental change in knowledge-intensive processes and in the related institutions. Some of this change is already evident in web sites that are dedicated to scientific knowledge. The extent of change between the traditional journal-based publishing system and the future online-based scientific communications system, however, is not clear. This article aims to provide insights about the extent of potential change.

Besides analyzing the impact of information technology this article also strives to demonstrate some of its possible implications. One of the key changes induced by information technology is the ability to provide multiple and customized presentations of an article. In the HTML version of this article, therefore, readers may select among different levels of detail and different perspectives depending upon their own interests. The sections selected are assembled dynamically by the browser, using JavaScript embedded in the web page. The text consists of more than 200 sections that can be assembled in more than 20 ways. The number of variations can be increased easily by adding more perspectives. For example, the reader can view text as:

- an approximately 150 word abstract,
- a 650 word extended outline,
- a 4000 word main article or
- a 7000 word in-depth document.

The creation of such adaptive texts will become much simpler as XML (Extended Markup Language) gains widespread support and more mature authoring systems for XML emerge. For the current article, the reader needs to specify her preferences manually. It is only a small step, however, to store user preferences in the browser (e.g. via browser cookies) and to assemble the document automatically to these preferences. The custom assembly of a document from individual sections has the advantage that ultimately reviewers' or readers' comments could be embedded within the text and made accessible on demand.

To illustrate the segmentation of the text, the sections in this text are color-coded. Text that appears in the abstract, extended abstract or main text is shown in the default color. Text - such as this section - which describes how to use this document, is shown in **red**. Additional detail which explains a theoretical or methodological issue is shown in blue; sections targeted towards application and practice are green; sections with further detail are marked in dark red. Sections may be relevant to several perspectives; then the most appropriate color is shown. The constitution of the whole text can best be viewed by selecting 'Detailed article' from the drop-down list box in the top frame and checking all check-boxes. In this document, color-coding only serves to illustrate the structure of the text. A more mature environment would combine sections from different documents into a custom text and should leave it entirely to the user whether or how to use such mark-up schemes. The reader might also profit if symbols could be included which indicate hidden information that could then be displayed on demand.

The article also provides direct access to the components of knowledge conveyed in this article. Figure 1 summarizes the contributions of this article around the main types of scientific knowledge. The classification of knowledge types is based on a meta structure of scientific knowledge which is developed in a separate paper (Hars 1999b). The meta structure consists of problems, concepts, theories, methods, data, statements, frameworks and several other components. Figure 1 lists the key contributions for each type of scientific knowledge. For example, as described in Section I., this article deals with three main problems:

- the impact of IT on the scientific communication system,
- alternatives for the design of scientific knowledge infrastructures and
- characteristics of scientific knowledge infrastructures.

The figure also lists the concepts that are relevant for online scientific knowledge infrastructures and describes the data on which the insights are based. The identification of different types of scientific knowledge contained in an article is a prerequisite for increasing the integration of scientific knowledge. If articles explicitly identified their individual knowledge contributions, then search engines could extract, collect, and integrate related contributions. For example, search engines could build a list of descriptions of concepts such as 'electronic journal' or 'online scientific knowledge infrastructure', which currently are deeply embedded within an article. This list could then be used as the basis for further scientific work about these concepts. Figure 1 uses color-coding to differentiate among aspects where original contributions are made in this article (green) and already existing knowledge on which this article is based (blue). In the online version of this text, the elements of the figure are hyper-linked to the corresponding sections within the document.

II. TYPES OF KNOWLEDGE INFRASTRUCTURES

In recent years, many institutions and individual researchers founded web sites dedicated to scientific knowledge. As the following analysis will show, their approaches vary greatly. While some sites focus on the dissemination of knowledge, others seek to create and integrate knowledge. While some of the sites transpose the traditional publishing system to the Internet, others experiment with novel approaches for managing scientific knowledge and aim to eliminate perceived deficiencies of the current system.

The current publishing system is deeply ingrained in scientific processes. It is not only based on technical innovations such as the printing press but also on institutional elements which evolved over time such as publishers, editorial boards, libraries, and tenure committees. The system can be used in many different ways and is highly flexible. It is taken for granted by many researchers. Thus the current publishing system has the characteristics of an infrastructure – a substrate on which research and scientific communication can be performed

(Star & Ruhlehder 1996, p.113). IT is changing this substrate and we are now witnessing the first elements of tomorrow's online knowledge infrastructures. In the online world, the term 'publishing infrastructure' may be misleading, however, because online infrastructures need not be limited to publishing. They may add support for the earlier phases of knowledge creation and integration. Therefore the term 'scientific knowledge infrastructure' will be preferred in this article.

The evolution of knowledge infrastructures has received much attention in the last fifty years. In 1945, Bush sketched a vision of a worldwide infrastructure which would allow the individual to access all human knowledge and which would support associative thinking (Bush 1945). Inspired by Bush's vision, Engelbart (1962) developed the first Hypertext system, Augment. Nelson worked on similar approaches and coined the term 'hypertext' for systems that allow "non-sequential writing" (Nelson 1981).

Hypertext became a synonym for a vision of future knowledge infrastructures in which fragments of texts are intensively cross-referenced in combination with a user interface which allows accessing text non-linearly (Spring 1991). The vision is strongly based on the goal of improving the interaction between human and computer and on reflecting the associative structure of human memory in the system. Because of the broadness of the visions, the implementation details of hypertext systems vary considerably. This ongoing stream of research resulted in many experimental hypertext systems as described by Münz (1997) and Conklin (1987).

Many characteristics of hypertext were incorporated into the World Wide Web. However, the current WWW still lacks support for key hypertext features such as support for joint authoring, versioning, and alternates. With maturation of the Internet, the first online journals began to emerge. At the end of the Eighties, Ginsparg, a scientist at Los Alamos National Laboratory, started a preprint archive that disseminated preprints via email. Today, this archive is one of the most successful online scientific knowledge infrastructures. In the beginning of the1990's, Harnad founded Psychologuy, an electronic journal, which allowed what he termed scholarly sky-writing – short scholarly pieces which are published immediately upon receipt and which invite rapid comment from peers (Harnad 1990).

RESEARCH METHOD

This study uses an exploratory approach. Online scientific knowledge infrastructures are rapidly evolving and not yet mature. The boundaries between the different types of knowledge infrastructures typically distinguished in the literature are fluid. Although meta-indexes exist for online scientific journals (for example http://www.coalliance.org/) there is a lack of indexes for other types of scientific infrastructures. Therefore this study focuses on the discovery of exemplars of different types of online knowledge infrastructures and the description and analysis of their characteristics. The types of knowledge infrastructures were determined based upon a search of the literature and an analysis of existing scientific online knowledge infrastructures. With the exception of the terms 'logic server', 'originally electronic journal' and 'hybrid electronic journal', all terms are mentioned in the literature. Knowledge infrastructures were identified by one of the following means:

- Analysis of existing web indexes.
- Querying search engines for terms such as 'collaboratory', or 'encyclopedia'.
- Using references from the scientific literature and following online-references.
- Exploring the World Wide Web.

A variety of online knowledge infrastructures exist. Figure 2 shows the main categories that are discussed in this section. The subsequent analysis will show that the boundaries between these infrastructures are fluid. Table 1 summarizes the characteristics of the categories and provides links to several exemplars for each category.

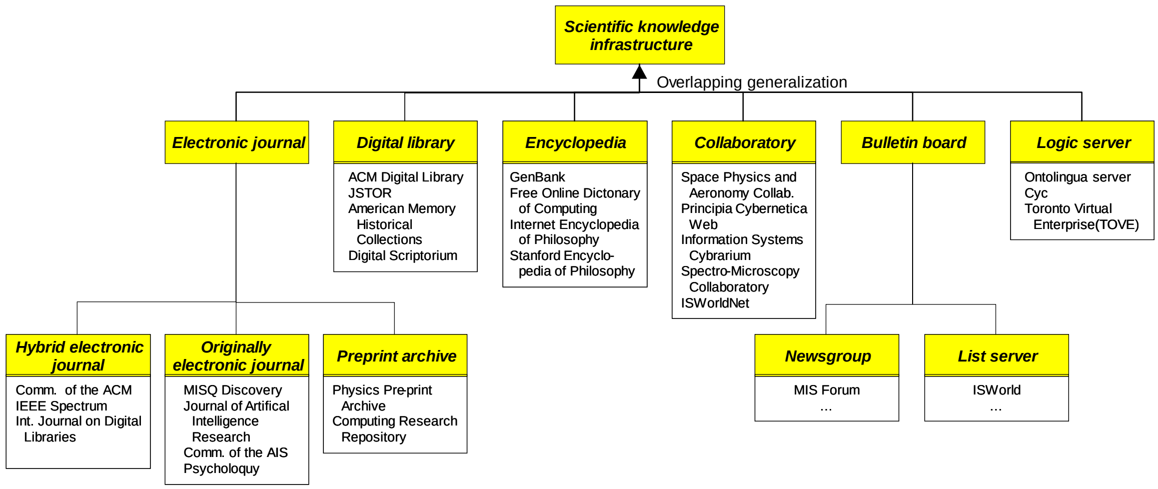


Figure 2: Types of Knowledge Infrastructures

Table 1: Types of Scientific Knowledge Infrastructures

Type		Characteristics	Comments	Examples
Electronic Journal		Focus on dissemination Articles are published after peer review Establishes record of authorship for scientific ideas Organized temporally into issues, volumes	Modeled after traditional journals	(see subcategories below)
	Hybrid electronic journal	Originated as paper journal Provides content of paper journal online	Most publishers are currently making their paper journals available online.	http://www.acm.org/cacm http://www.spectrum.ieee.org http://link.springer.de/link/service/journals/00799/index.htm
	Originally electronic journal	Originated as electronic journal Organized temporally but may have eliminated issues	May have a paper presence	http://www.jair.org http://cais.isworld.org/ http://www.misq.org/discovery
	Preprint archive	Rapidly disseminates unreviewed articles May have additional non-temporal organizing criteria (by topic, by author etc.) Instant publishing of submitted articles without review	Preprint archives disseminate articles rapidly without review (most preprints are simultaneously submitted to a scientific journal for review)	http://xxx.lanl.gov http://xxx.lanl.gov/archive/cs

Digital library	Aggregation of knowledge resources (e.g. journals, encyclopedias etc.) Focus on dissemination	May consist of any combination of knowledge infrastructures	http://www.jstor.org http://www.acm.org/dl http://memory.loc.gov/ammem/ammemhome.html http://sunsite.berkeley.edu/scriptorium/form.html
Encyclopedia	Extensively cross-referenced knowledge Focus on consistency of knowledge Multiple entry points and reading paths Organized alphabetically or by topic (not temporally) Frequently keeps no record of authorship	In contrast to journals focuses on the relationships between different contributions, emphasizes the organization of knowledge.	http://www.ncbi.nlm.nih.gov/Web/Genbank http://wombat.doc.ic.ac.uk http://www.utm.edu/research/iep http://plato.stanford.edu/
Collaboratory	Focus on communication between scientists May provide joint access to instruments Focuses on knowledge creation, integration and dissemination Contributions are published instantly	Originally viewed as a 'virtual laboratory' where researchers have joint access to instruments, the term is being increasingly used for highly interactive knowledge infrastructures.	http://intel.si.umich.edu/SPARC http://pespmc1.vub.ac.be http://cybrarium.usc.edu http://www-itg.lbl.gov/BL7Collab/

Bulletin board	Focus on communication between scientists Contributions are published instantly Organized temporally by date/time of contribution	Ubiquitous on the Internet.	(see subcategories below)
News group	Uses a pull approach for disseminating knowledge		http://cism.bus.utexas.edu/issues/iindex.html
List server	Uses a push approach for disseminating knowledge		http://www.commerce.uq.edu.au/isworld
Logic server	Formal representation of knowledge Support for online creation and editing of knowledge Mechanisms for consistency checking Focus on reuse of knowledge	Logic servers have been pioneered in the AI community.	http://www-ksl-svc.stanford.edu:5915 http://www.cyc.com/cyc-2-1/toc.html

ONLINE JOURNAL

Online journals are one of the predominant types of web sites dedicated to scientific knowledge. They replicate core characteristics of traditional, paper-based journals in the online environment. Thus online (or electronic-) journals focus on the dissemination of knowledge in the form of articles. As in paper journals, articles are largely self-contained units of knowledge that can be compiled into issues and volumes. Thus, although information technology eliminates the need for collating articles by publishing date, most online journals retain time as primary ordering criterion and primary access path. Like traditional journals, online journals are concerned with the authorship of ideas and the associated priority claims. To ensure quality, online journals typically implement a peer review process. While it may be modeled on the traditional review process, some journals experiment with innovations such as providing workflow support for refereeing. Workflow support allows keeping track of each concern raised by reviewers as well as the authors' response. It may thus lead to more detailed and systematic 'micro-reviews' during which reviewers and authors interact more intensively (Barua, Chellappa, and Whinston, 1997).

Many journals that are currently available in electronic form originated in the paper environment and continue their main presence in paper form. Because such journals remain strongly bound to traditional structures they will be called 'hybrid online journals'. Most major publishers of print journals (e.g. Springer and Elsevier, as well as scientific associations such as IEEE and ACM) established or are currently establishing electronic versions of their traditional journals. IEEE and ACM already provide all current articles in their print journals in online form. An interesting aspect of hybrid journals is the time at which an accepted article is published electronically. While most hybrid journals publish the online version at the same time as the print issue, some journals adopt a 'preview' policy where the online version of an article is published as soon as it is accepted. Previews may appear before the other articles to be published in an issue are finalized. This phenomenon is an indication that a necessary characteristic of traditional journals – that articles are bundled into issues – no longer has the same status in an electronic environment.

A smaller number of online journals were developed directly for the online environment. Examples are the Journal of Artificial Intelligence Research (JAIR) and the Communications of the Association for Information Systems (CAIS). These journals, which will be called 'originally electronic journals', are not bound by the need to publish a print version, although some opt to issue a print or CD-Rom version of each volume.

This is the case at JAIR which publishes a paper volume at the end of each year. Although the articles can be accessed free of charge on the World Wide Web, there seems to be sufficient demand for the more costly paper version by libraries. Some originally online journals, such as CAIS, have implemented additional changes. They not only publish each article as it is accepted but completely eliminate the bundling of articles into issues. While most originally electronic journals provide access via a traditional table of contents that orders articles by publishing date, some journals use different approaches for structuring the contents. The JAIR, for example, provides an alternative view, the 'JAIR information space', where articles are organized by topic.

A screen shot of JAIR is shown in Figure 3. The information space consists of three sections. One section (to the right) displays all articles grouped by topic. The distance between the circles for each topic represents the conceptual distance between the topics. On the top left, the title and abstract of the currently selected article are shown. The bottom left is a tabbed section which allows to list articles by author or titles and which provides access to the search engine. JAIR demonstrates one key advantage of an online environment: the ability to link articles into a dynamic knowledge structure which provides direct access to all articles concerning a specific topic independently of their publishing date.

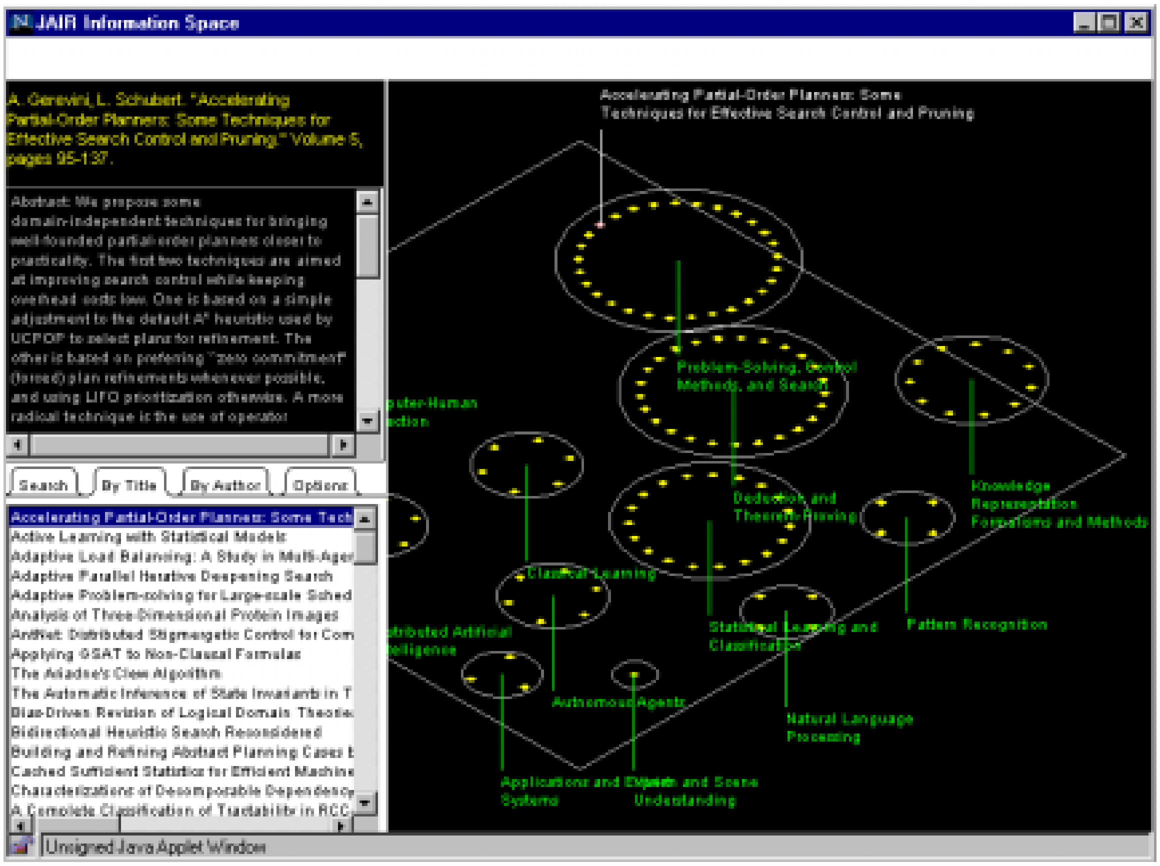


Figure 3. [JAIR Information Space]() (http://jair.org)

Pre-print journals are another variant of online journals. Pre-prints are copies of unpublished manuscripts submitted to a scientific journal for review. This definition is prototypical: although most current preprints meet the criterion of being targeted at a journal, working papers may occasionally also be distributed as preprints. Authors in disciplines such as physics and biology have a long tradition of exchanging preprints to rapidly alert their peers about new developments. In the 1970's and 1980's many Universities operated their own preprint presses specialized for this purpose. Pre-prints drastically reduce the delays of the traditional publishing system without eliminating the rewards of being published in reputed journals.

At the end of the Eighties, Paul Ginsparg, a physicist at Los Alamos, realized that the Internet would be an ideal vehicle for the distribution of preprints. He then started an electronic preprint service which was modeled on a list server. As expected, the Los Alamos Preprint Archive greatly simplified the dissemination of preprints. It was an instant success and since has been extended to several other areas including computing research. In September 1999 the archive sustains more than 100.000 hits per day and is replicated in 14 countries. As the World Wide Web gained in popularity, a web-based interface was added to the archive. It allows authors to upload their preprints in several different formats. Each preprint can be assigned to one or several topical areas. Readers can then download preprints in those areas and time periods in which they are interested. The interface is very basic and the underlying technology is simple. Administration requirements and costs of operation are minimal. This archive shows the value of knowledge infrastructures that concentrate on the core advantages of the system - in this case the instant dissemination of scientific articles to anyone interested - rather than on bells and whistles. Preprint journals differ from traditional journals in one essential aspect: articles are accepted without review and published instantly for retrieval by interested readers.

DIGITAL LIBRARY

Digital libraries aim to bring the library to the user (Story et al 1992). They provide electronic access to large bodies of knowledge. Many publishers are currently developing digital libraries to allow access to their journals through a uniform interface. Digital libraries frequently are aggregations of other online resources such as online journals, encyclopedias, and collaboratories. ACM, for example, provides access to all recent articles of all its journals in the ACM Digital Library. While any reader has access to tables of contents and abstracts, only digital library subscribers can access the whole content. However, non-subscribers can purchase access to individual articles.

Many university libraries subscribe to such online libraries and then redistribute access to their faculty and students. Other approaches to digital libraries aim to develop collections of knowledge for specific topics. An example is the American Memory Historical Collection maintained by the Library of Congress. It provides free access to more than 50 historical online resources via the web. The digital library thus links and integrates heterogeneous knowledge which is distributed across many sites.

Another example of a digital library is JSTOR (short for Journal Storage Project), an archive of digitized articles from print journals. JSTOR is unique because it provides access only to journal articles that are older than three years. This policy significantly reduces the costs for the digital library. Because the main revenue stream for journal publishers occurs at the time an article is first published, JSTOR can purchase access rights without harming the publisher's profitability. JSTOR now provides access to more than 100 journals in 15 disciplines and counts a large number of university libraries among its subscribers. Thus JSTOR is an example of how the needs of an established traditional system can be preserved while at the same time exploiting the new potential of technology.

ENCYCLOPEDIA

In the 18th century, an alternative to topical books and learned journals appeared in France and England (Guedon 1996, p.77). Encyclopedias aim to collect the core knowledge that is relevant in a field and organize it for rapid access by readers who may have very different needs and interests. As a consequence, the way in which knowledge is structured becomes a core concern. Encyclopedias use cross-references extensively to allow the reader to find his unique path through the knowledge rather than forcing each reader on the same path. Thus encyclopedias differ significantly from the journal paradigm. While authorship is very important in journals, the concept is less central to encyclopedias. When many persons contribute to an entry and when entries are updated frequently, it becomes difficult to clearly attribute the role of each participant. When knowledge is synthesized from many sources, when only a few sentences are changed from a previous version, who is the author? Encyclopedias show that much of the body of scientific knowledge is a collective enterprise that is based on many small contributions of a large number of individuals.

Many knowledge infrastructures in encyclopedic form are available on the Internet. An example from the field of biology is GenBank, a database of genome data (Burks et al. 1990). Authors provide the results of their gene sequencing efforts and link them into a common taxonomy. Contributions to GenBank are highly structured around predefined fields to ensure compatibility between entries. GenBank provides additional functionality such as custom editors and algorithms for detecting similarities in gene sequences.

Another example are encyclopedias in philosophy. Both the Stanford Encyclopedia of Philosophy and the Internet Encyclopedia of Philosophy are organized around the key topics of philosophy. Contributors from many areas of philosophy are working together to provide a summary of current knowledge in the field. In the case of the Stanford Encyclopedia, contributors are able to update their contribution at any time. Any change is immediately reflected in the encyclopedia. However, the encyclopedia notifies reviewers about changes and thus provides an additional mechanism for quality assurance. In contrast to traditional journals, however, notification occurs after publication. In an online environment, where published knowledge can easily be retracted, this approach preserves the ability to incorporate changes quickly without compromising quality.

Another example of an encyclopedia is the Free Online Dictionary of Computing (FOLDOC). It contains brief explanations of computing terms that are intensively cross-referenced. The encyclopedia is tightly controlled by an editor but is based on the contributions of more than 800 authors. Unlike the practice in journals, authors are only listed for the encyclopedia as a whole rather than for each individual entry.

In the paper environment, the differences between journals and encyclopedias are hard to bridge. In paper journals, cross-references cannot be introduced after an article is published and articles cannot be grouped together with other articles about the same topic published earlier. In an online environment, these boundaries can be removed. As the Journal of Artificial Intelligence Research

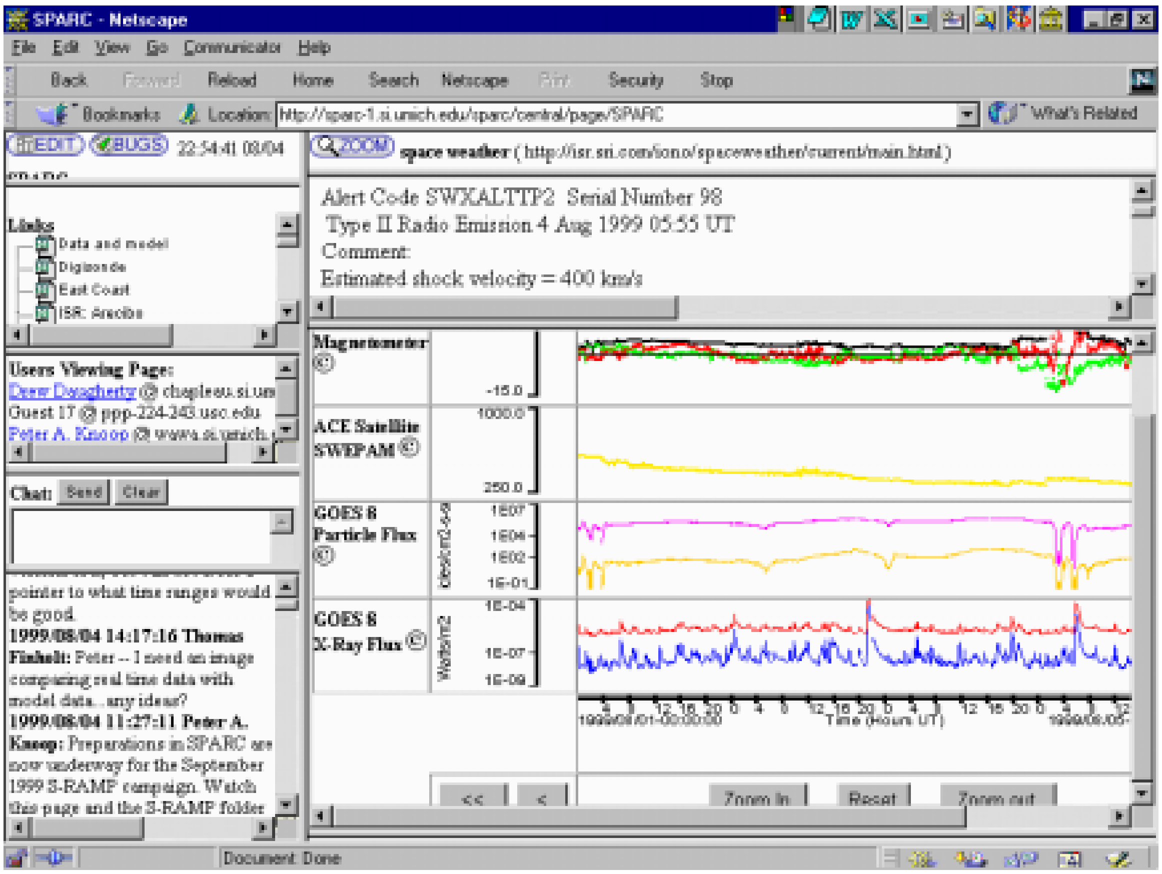
(JAIR) shows, journals can overlay a topical structure on top of the individual articles. In addition, links and cross-references between articles can be introduced technically even after an article has been published. Thus online infrastructures may exhibit characteristics of journals and encyclopedias simultaneously.

COLLABORATORY

While journals focus on the dissemination of knowledge and encyclopedias focus on the structuring of knowledge, collaboratories emphasize the creation of knowledge. Collaboratories bring researchers together in a virtual space where they can jointly perform research, access instruments, and discuss their results. The term collaboratory was coined in the late 1980's in the context of an initiative of the National Science Foundation which aimed to build a 'research center without walls in which the nation's researchers can perform their research without regard to geographical locations – interacting with colleagues, accessing instrumentation, sharing data and computational resources [and] accessing data in digital libraries' (Wulf 1989).

One of the first collaboratory projects was the Upper Atmospheric Collaboratory, renamed the Space Physics and Aeronomy Collaboratory (SPARC). It allows researchers to collaborate by pooling the instruments of several globally dispersed research groups, for examples satellites and radars. In addition, it provides access to atmospheric models and supports discussions. SPARC is frequently used for 'campaigns' in which researchers from different parts of the world collaborate in jointly developed, well-planned experiments that last several days.

Figure 4 shows an example of the SPARC web interface. It consists of a collection of frames each of which can be resized by the user. The large window on the lower right shows real-time data from several atmospheric instruments. The user can zoom in and out. The middle window on the right contains textual information about recent atmospheric events; the topmost window on the right is minimized. It displays the current space weather report. The left hand side ('links') shows an tree of accessible resources most of which also provide access to real-time data in graphical form. The middle left contains a chat window. The chat text is displayed in the window at the bottom left. This infrastructure brings atmospheric researchers together in ways that were impossible previously.



.

Figure 4. SPARC Screenshot

Another example of a collaboratory is the Information Systems Cybrarium developed by the author. It focuses on the creation and integration of knowledge in information systems and allows authors to submit articles without review. Every reader can provide comments and evaluate each article according to several criteria. These evaluations are used as parts of filters that enable each reader to define their own preference function for articles of interest to them. For example, readers can choose among different levels of rigor, relevance, and detail in the articles. In addition, the Cybrarium distinguishes different types of scientific contributions. General articles, which contain a variety of insights are distinguished from more systematic contributions which describe a scientific concept, theory or instrument. The Cybrarium also contains a separate category for describing scientific problems.

Several additional collaboratories are currently being developed with initial support from the National Science Foundation (Johnston & Sachs 1997). For example, a joint project between the Lawrence Livermore Laboratory and the University of Wisconsin-Milwaukee aims to provide access to experimental tools to nine participating institutions. A list of additional collaboratories is accessible at: http://www.techreview.com/articles/ma98/rflinks.html.

BULLETIN BOARD

A bulletin board is an electronic message database which allows users to read and post messages. Subcategories are listservers that push messages to readers and newsgroups that implement a pull strategy (Kendall and Kendall, 1999). These systems are one of the most successful knowledge infrastructures on the Internet. Their success lies in enabling the communication between individuals interested in a topic and the ability to learn from the communication of others. Besides answering specific questions, bulletin boards alert readers to novel problems and to issues they would not have thought asking about. In addition, many bulletin boards archive the discussions and provide search functionality. However, relevant knowledge becomes quickly drowned in large numbers of similar but irrelevant messages. As a consequence, the repository function of bulletin boards is much less important than their communication function. Bulletin boards show that enhancing communication must be a core function of each knowledge infrastructure.

This category of knowledge infrastructures differs greatly from those discussed previously. Logic servers are sites that allow researchers to describe scientific knowledge formally using a logic-based language. They allow the specification of ontologies – formal definitions of concepts that are relevant in a domain and of their attributes and relationships (Guarino & Giaretta 1995). Stanford's Ontolingua Server, for example, allows any user to author an ontology. It provides an editor for the definition of knowledge and various services for analyzing the content and ensuring that the knowledge is consistent (Farquhar, Fikes & Rice, 1996). In addition, Ontolingua Server allows reusing existing ontologies. Thus authors can import core ontologies, e.g. for describing key spatial and temporal concepts rather than defining them themselves. A simple time ontology, for example, defines concepts such as a 'point in time', 'interval', temporal relationships such as 'before', 'after' etc. Several attempts have been made to use logic servers for formalizing management knowledge. However, they are far from being complete and it is questionable whether logic-based formalisms are suitable for the inherent complexity (and as a consequence imprecision) of the management domain. An example of a management-oriented ontology that uses Ontolingua Server is the Enterprise Ontology developed at the Artificial Intelligence Applications Institute at the University of Edinburgh (Uschold et al. 1998). The ontology currently consists of 92 core concepts related to an enterprise. The lower frame in Figure 5 shows the ontology's top-level classes. For each class, attributes and relationships are defined. The ontology is roughly divided into four sections: Activity and processes, organization, strategy and marketing.

Logic servers point out additional characteristics of knowledge infrastructures which may be relevant in a management context: the need for knowledge editors which allow entering knowledge and which simplify the identification and creation of relationships to other units of knowledge in the infrastructure. Another aspect is the algorithms for detecting inconsistencies between new and existing knowledge. Thus logic servers show additional ways of manipulating scientific knowledge.

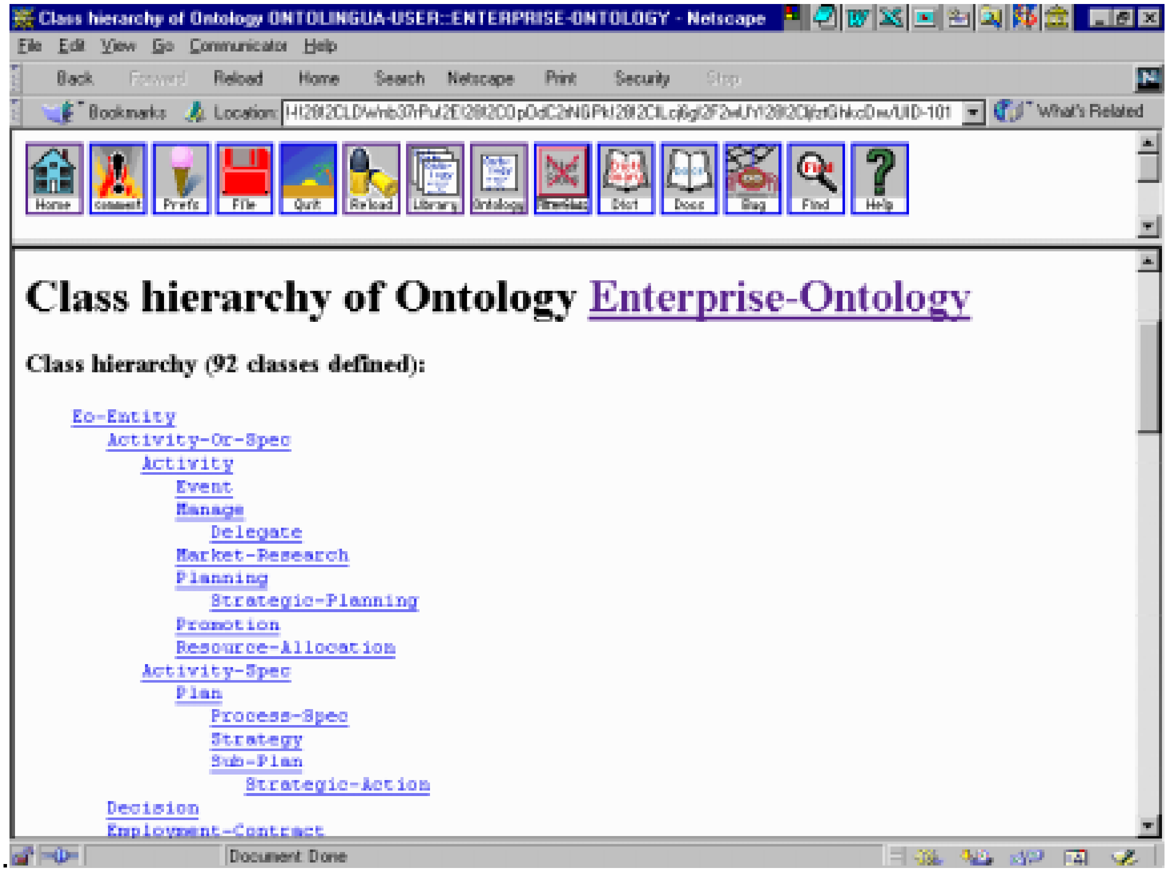


Figure 5: An Example Class Hierarchy (Enterprise Ontology) in Ontolingua Server

In the social sciences, custom editors may enable an author to systematically describe key attributes of a theory, a variable or an instrument. It may also allow to incorporate existing concepts, variables and instruments into a new theory. Similarly, automated algorithms may search a scientific knowledge infrastructures to find theories or methods that have substantial overlap in structure or the concepts used.

III. OVERVIEW OF CHARACTERISTICS

In Section II, many characteristics of online scientific knowledge infrastructures were identified. The most important among them are discussed in this section.

MODE OF INTERACTION

Online infrastructures differ in the way in which users interact with the system. In electronic journals, interaction is mostly limited to reading. Other infrastructures turn the focus towards the earlier phases of the research process and provide support for editing, annotation, and comments. The mode of interaction is one of the key criteria for the design of online knowledge infrastructures. While the nature of paper limited traditional journals to focus on dissemination, online infrastructures can increase the level of interaction. This increase may occur gradually. A journal may initially support comments and annotation and only later turn towards automated support for the review, annotation, creation, and integration of contributions.

TIME TO PUBLISH

Several infrastructures depart from the traditional sequence of authoring, review and publishing. For example, preprint journals, bulletin boards and collaboratories publish contributions as soon as they are submitted and before they are reviewed. Thus the designers of a knowledge infrastructure need to decide upon the desired publishing delay. Instant publishing of an article submitted to a knowledge infrastructure reduces publishing delay effectively to zero and increases communication among peers. The disadvantage is that some control over article quality is lost. However, in contrast to the print environment, in an online world publishing is reversible. If the review is negative, the article is simply removed from the site. This approach is already visible in some journals which provide a separate space for articles which are under review. In an online environment, the concept of 'separate' spaces, however, is misleading. It assumes that the infrastructure is static and that each reader is presented with the same layout and content. When the presentation is customized automatically to the preferences of each reader, however, the structure of the publishing space may differ from reader to reader. While some readers may use the distinctions among accepted, in review, and rejected papers as their primary criterion for structuring their space, others may prefer to access articles by topic and include rejected papers in some topics. This arrangement may not only be useful for maintaining the history of work but also allows researchers to stay abreast of the activities of direct competitors who perform similar research.

REVIEW

The online environment can change the review process in several ways. Workflow software speeds up the process and allows tracking the activities of readers and reviewers. More important, review no longer needs to be binary. Instead of summarily evaluating an article, reviewers may evaluate it along several dimensions, such as rigor, relevance, and originality. While journals may aggregate these criteria for their acceptance decision, the criteria can also be used by readers to create personalized review functions. When each article contains these evaluations, then tables of contents can be generated dynamically depending on the individual reader's preferences for criteria such as rigor, relevance and originality. This arrangement improves the chances for acceptance of 'niche articles' of interest only to small audiences.

The online environment can also support mechanisms other than traditional peer review. Examples are reader evaluations, recommendations based on access or usage statistics, and collaborative filtering approaches. Collaborative filtering is a technology that generates recommendations for users based on others' evaluations (Konstan et al. 1997). A collaborative filtering system records the preferences or evaluations of its users. Based on an analysis of the preferences it generates a profile of each user. When a user looks for new items he would be interested in, the system identifies users with a similar profile and then recommends items they ranked highly.

STRUCTURE OF KNOWLEDGE

The infrastructures differ in the way in which they represent and organize knowledge. In journals, the article is the core unit of knowledge. Articles are collections of many individual scientific insights. Other infrastructures, in contrast Online infrastructures may use a more fine-grained unit of knowledge than traditional articles. Bulletin boards, are limited to a few sentences. Logic servers break knowledge down into small logic statements.

A second difference is the way in which knowledge is integrated and cross-referenced. Journal articles are sparsely linked through references. Readers are accustomed to many additional implicit interdependencies between an article and other published knowledge. Paper references also have the disadvantage that they point only in one direction.

A third aspect is the way in which knowledge is organized. While some electronic journals provide access to journal articles through a table of contents - which is ordered by publication date - other infrastructures organize knowledge by subject area and intensively cross-reference knowledge. These approaches do not exclude one another. Therefore electronic journals may add subject area indices. In addition, they may integrate each new contribution into a higher level of structure of knowledge where articles are cross-referenced systematically.

AUTHORSHIP

Journals not only disseminate knowledge, they also record the evolution of knowledge and establish the authorship of novel scientific ideas. Authorship plays a central role in the current scientific communication system. However authorship was not always that important. In the Middle Ages authors were frequently not listed. Encyclopedias do not mention the authors of an entry. While it may be possible to determine the author of a new scientific idea, it is much more difficult to attribute authorship when knowledge is synthesized from many different sources and when many authors collaborate on a piece of knowledge. Thus, the more integrated a body of scientific knowledge becomes, the more the generation of scientific knowledge becomes a collective enterprise, and the more difficult it becomes to determine the author. Of course, it is technically possible to record the authorship for every paragraph, sentence, and word. However, the problem lies in the aggregation of authorship. Who is to be listed at the head of an article that went through many iterations and refinement over the years? Should the original author remain the 'first author' - even if few of his original sentences remain? Does it discourage current scientists from keeping the knowledge up-to date? Do we need different perspectives on authorship? It is not clear what the answers will be.

IV. IMPLICATIONS

The preceding survey showed that knowledge infrastructures adopt a wide range of approaches. While the majority of current infrastructures are still modeled on the traditional journal, the other infrastructures point to many implications for the evolution of scientific knowledge infrastructures. The most important implications are briefly summarized in this section.

FROM DISSEMINATION TO COLLABORATIVE CREATION OF KNOWLEDGE

In the paper environment, articles are static; the primary focus is on the dissemination of knowledge. Direct communication between reader and author or between different authors is not possible. In the online environment, however, articles can be updated and links and annotations can be added easily. Thus, knowledge becomes dynamic and knowledge infrastructures can change the focus towards supporting the processes of knowledge creation and integration, which includes discourse among authors and between author and reader. Thus online infrastructures allow a more complete support of the scientific research process. They are not limited to dissemination. In the long run this change may lead to a reassessment of the finality of scientific knowledge. While paper articles led to a concept of scientific insights being 'etched in stone' (Harnad 1992), information technology may lead to a different understanding of scientific knowledge – as sparks in a continuous scientific discourse.

MORE FOCUS ON THE STRUCTURING AND INTEGRATION OF KNOWLEDGE

Information technology allows embedding contributions in a dynamic web of relationships. Therefore the structuring of knowledge and the integration of insights into the existing body of knowledge will become more important. In the paper environment, category systems and references cannot lead the reader directly from one article to another. In the online environment, in contrast, knowledge structures can be maintained in parallel with individual contributions and each contribution can be linked into the structure so that readers can follow the links and access the linked contributions directly. This change is important for references and citation. In the paper environment, these links only work in one direction: the reader of a cited article cannot determine which articles cite the work being read. Thus the evolution of knowledge can be traced backward, but not forward. Information technology allows bi-directional links. When citations are stored in a database, it is possible to find out which articles have cited a particular work. Unfortunately, the current implementation of the World Wide Web is based on unidirectional hyperlinks. This limitation does not prevent knowledge infrastructures which are built on top of the WWW to implement a bi-directional link mechanism.

FLUID BOUNDARIES BETWEEN DIFFERENT TYPES OF KNOWLEDGE INFRASTRUCTURES

Although the types of knowledge infrastructures discussed in this article differ in many characteristics, the boundaries of each type of infrastructure are not fixed. An electronic journal, for example, may incorporate interaction features such as reader annotation or collaborative editing. These features move the journal towards a collaboratory. The journal may also add detailed subject area indices and implement an alternative encyclopedic access structure to the articles. It then becomes increasingly difficult to determine to which of the knowledge infrastructure categories the infrastructure belongs. Similar to Internet search engines (e.g. Altavista) and Internet directory services (e.g. Yahoo) which originated as clearly separate categories but now have adopted many of each others characteristics, the evolution of online scientific knowledge infrastructures may lead to the dissolution of the currently still evident boundaries among the different categories discussed above. This does not imply that all future knowledge infrastructures will be identical. However, the overlap in their characteristics will increase greatly.

ONLINE DOCUMENTS MAY ADAPT TO THE USER

While traditional articles are static, the presentation of online documents can be adapted to the specific needs of each reader. Documents may be assembled automatically from many small segments of scientific knowledge. These segments need not all be part of a single document but may come from more than one article. Thus texts may change as new knowledge is added to an online knowledge base. A key issue for using adaptive documents is the further maturing of XML and creation of standards for marking up elements of scientific knowledge within a document.

V. CONCLUSION

The transition towards mature online knowledge infrastructures poses many scientific and practical challenges. The analysis demonstrates that:

1. Information systems researchers need to develop novel computer-supported approaches for structuring, organizing, integrating, and visualizing scientific knowledge.

2. The scientific community needs to rethink its institutions for scientific communication and restructure its reward system.

Change will not be easy or quick. It takes time to unlearn conventional wisdoms e.g. that articles need to be reviewed before they are published or that scientific insight is best packaged in articles that consist of hundreds of scientific statements. Furthermore, revenue models need to be adapted.

Information systems researchers have long argued businesses need to apply IT to reinvent and transform their processes. It is time to apply this insight to our own domain.

Editor's Note: This article was received on September 6, 1999. It was with the author for three and a half months for three revisions. The article was published on July 27, 2000

REFERENCES

EDITOR'S NOTE: The following reference list contains hyperlinks to World Wide Web pages. Readers who have the ability to access the Web directly from their word processor or are reading the paper on the Web, can gain direct access to these linked references. Readers are warned, however, that

1. these links existed as of the date of publication but are not guaranteed to be working thereafter.
2. the contents of Web pages may change over time. Where version information is provided in the References, different versions may not contain the information or the conclusions referenced.
3. the authors of the Web pages, not CAIS, are responsible for the accuracy of their content.
4. the author(s) of this article, not CAIS, is (are) responsible for the accuracy of the URL and version information.

Barua, A., Chellappa, R., and Whinston, A.B. (1997), "Creating a collaboratory in cyberspace: Theoretical foundation and implementation". In: *Journal of Organizational Computing*, 5(4), pp. 417-442.

Burks, C. et al. (1990), "GenBank: Current status and future directions". In: Doolittle, R.F. (Ed.): Molecular evolution: Computer analysis of protein and nucleic acid sequences. *Methods in enzymology*, vol. 183, pp. 3-22.

Bush, V. (1945), "As we may think". *Atlantic Monthly*, 176(1), pp. 101-108, http://www.theatlantic.com/unbound/flashbks/computer/bushf.htm.

Conklin, J. (1987), "Hypertext: An introduction and survey". *IEEE Computer*, 20(9), pp. 17-40.

Engelbart, D.C. (1962), *Augmenting Human Intellect: A Conceptual Framework*. Summary Report Contract AF 49(638)-1024, Menlo Park, CA: Stanford Research Institute, October 1962.

Farquhar, A., Fikes, R., and Rice, J. (1996), The Ontolingua Server: A tool for collaborative ontology construction. : Stanford Knowledge Systems Laboratory

Report KSL-96-26.Stanford, CA: Stanford University, ftp://ftp-ksl.stanford.edu/pub/KSL_Reports/KSL-96-26.ps.

Guarino, N., and Giaretta, P. (1995), "Ontologies and knowledge bases - Towards a terminological clarification." in Mars, N.J. (ed.): *Towards very large knowledge base - Knowledge bases and knowledge sharing*, Amsterdam:IOS Press, pp.25-32.

Guedon, J.-C. (1996), "The seminar, the encyclopedia, and the eco-museum as possible future forms of electronic publishing." in Peek, R.P. Newby, G.B. (eds.): *Scholarly publishing - The electronic frontier*. Cambridge (MIT Press), pp. 71-89.

Harnad, S. (1990), "Scholarly skywriting and the prepublication continuum of scientific inquiry." *Psychological Science* 1, pp. 342 - 343, http://www.cogsci.soton.ac.uk/~harnad/Papers/Harnad/harnad90.skywriting.html.

Harnad, S. (1992), "Interactive Publication: Extending American Physical Society's Discipline-Specific Model for Electronic Publishing." *Serials Review, Special Issue on Economics Models for Electronic Publishing*, pp. 58 - 61.

Hars, A. (1999a), A theory of scientific online journals: The impact of IT on the publishing process. IOM Department Working Paper. Los Angeles, CA: University of Southern California, Marshall School of Business, http://cybrarium.usc.edu/cyb.dll/show?object=1001950.

Hars, A. (1999b), Towards epistemologically adequate representation of scientific knowledge. IOM Department Working Paper. Los Angeles, CA: University of Southern California, Marshall School of Business, http://cybrarium.usc.edu/cyb.dll/show?object=1002000 .

Kendall, J.E. and Kendall, K.E. (1999), "Information delivery systems: An exploration of web push and pull technologies." *Communications of the Association for Information Systems 1*(14), http://cais.aisnet.org/articles/default.asp?vol=1&art=14.

Konstan, J.A. et al. (1997), "GroupLens: Applying collaborative filtering to Usenet news." *Communications of the ACM*, 40(3), pp. 77-87.

Johnston, W.E and Sachs, S. (1997), Distributed, Collaboratory Experiment Environments (DCEE) Program: Overview and Final Report, February, 1997. Berkeley, CA: Lawrence Berkeley National Laboratory, http://www-itg.lbl.gov/DCEEpage/DCEE_ Ov erview.html (8/2/1999).

Münz, S. (1997): Hypertext. V.1.1. Zuerich: NetSystems, http://www.netsystems.ch/support/hypertext/htxt.htm (8/2/1999).

Nelson, T. (1981) Literary machines. Swathmore (self-published), http://www.sfc.keio.ac.jp/~ted/TN/PUBS/LM/LMpage.html (7/6/2000).

Spring, M.B. (1991) *Electronic printing and publishing*. New York: Marcel Dekker.

Star, S.L. and Ruhleder, K. (1996), "Steps toward an ecology of infrastructure: design and access for large information spaces." *Information Systems Research*, 7(1), pp. 111-134.

Story, G.A., O'Gorman, L., Fox, D., Levy Schaper, L., and Jagadish, H.V. (1992), "The RightPages image-based electronic library for alerting and browsing." *IEEE Computer*, 25(9), pp. 17-26.

Uschold, M., King, M., Moralee, S., and Zorgious, Y. (1998), "The Enterprise Ontology." *Knowledge Engineering Review*, Vol.13, ftp://ftp.aiai.ed.ac.uk/pub /documents/1998/98-ker-ent-ontology.ps.

Wulf, W. (1989), "The national collaboratory - A whitepaper". in Lederberg, J. and Uncapher, K. (eds.): *Towards a national collaboratory: Report of an invitational workshop at the Rockefeller University.* Washington, DC: NSF, Directorate for Computer and Information Science and Engineering, Appendix A.

ABOUT THE AUTHOR

Alexander Hars is Assistant Professor of Information Systems at the Marshall School of Business at the University of Southern California. His research interests include knowledge management, enterprise modeling and the application of natural language processing to systems analysis and design. He is the founder of an internet based scientific knowledge infrastructure, the IS Cybrarium. He has published in Communications of the ACM, Journal of Database Management, Information Management etc. He obtained his doctorate degree from the University of the Saarland (Germany).

Communications of the Association for Information Systems