1. Leveraging information technology for science

1.1. Motivation

Management theory increasingly recognizes the importance of knowledge as an additional production factor besides the classical factors (Gutenberg 1965) of labor, equipment and materials (Grant 1996; Kogut & Zander 1992; Nelson 1991; Drucker 1998). In the transition towards a knowledge based economy, information technology plays a crucial role. Thus many firms and institutions are currently reengineering their knowledge-intensive processes and are building IT-based knowledge infrastructures. Functional organization structures are being replaced by process-oriented structures. Processes are not only being shortened by reducing transfer times but also by parallelizing previously sequential tasks. Access to information and decision-making authority is being decentralized (while establishing centralized monitoring and exception handling mechanisms). In many cases, the concept of 'product' has been refocused around knowledge-intensive services. Moreover, many efforts are targeting the collection and organization of knowledge. Consulting companies, for example, are pioneering knowledge bases which organize and replicate knowledge about best practices (for an overview see Prusak 1998).

One of the areas where novel approaches for the management of knowledge have not yet taken a deeper hold are the sciences. Although they are very knowledge-intensive, their decentralized organization and the high complexity of scientific knowledge have been obstacles in the move towards more efficient management of scientific knowledge.

Many voices within the scientific community have recognized these problems and argued that better mechanisms for the accumulation and utilization of scientific knowledge are needed:

> "There is a growing mountain of research. But there is increased evidence that we are being bogged down today as specialization extends. The investigator is staggered by the findings and conclusions of thousands of other workers -- conclusions which he cannot find time to grasp, much less to remember, as they appear. Yet specialization becomes increasingly necessary for progress, and the effort to bridge between disciplines is correspondingly superficial. Professionally our methods of transmitting and reviewing the results of research are generations old and by now are totally inadequate for their purpose." ..."The difficulty seems to be, not so much that we publish unduly in view of the extent and

variety of present day interests, but rather that publication has been extended far beyond our present ability to make real use of the record." (Bush 1945)

Information technology provides many opportunities for addressing these problems. Many online scientific journals have been established which facilitate access to scientific knowledge, reduce publishing cost and shorten processing times for review and dissemination. However, as Hammer and Champy (1994) have argued, information technology must lead to more fundamental change than automating and accelerating traditional processes. Thus the capabilities of information technology may fundamentally change the scientific publishing system and the way in which scientists create, organize and disseminate knowledge. Such change is already visible in some online infrastructures, most notably in physics (Ginsparg 1994) and molecular biology (Benson et al. 1998), which differ in significant aspects from the traditional 'journal' model.

Nevertheless, most academic sites replicate traditional approaches which scientists have become accustomed to over the centuries. For example, articles continue to be treated as being immutable – although electronic articles could be updated easily. Many electronic journals still publish issues, although bundling electronic articles does not provide similar distribution advantages as bundling paper. Adhering to established structures is risky: reputable journals risk to be marginalized by nimble competitors that establish more effective sites. Moreover, better approaches to structuring and disseminating scientific knowledge may fundamentally change the way in which research is conducted and thus make research more productive and satisfying.

As the next two sections will show, the impact of new information technologies on the practice of science can hardly be overestimated. The next section discusses the close relationship between significant changes in information technology and scientific progress. To substantiate the argument, a brief analysis of potentials for change follows. This paves the way to refine the concept of sites which are dedicated to scientific knowledge ('scientific knowledge infrastructures') and to provide an overview over the subsequent chapters.

1.1.1. Impact of previous IT revolutions on science

The history of knowledge is inextricably linked with advances in information technology. As Table 1 shows, the first revolution in information technology was the invention of writing in Mesopotamia around 3500 B.C. (Jackson 1981, p 16; Van Doren 1991 p 10). It was an important factor for the emergence of the first sophisticated civilization. Writing allowed record-keeping and communication of rules and knowledge. Reading and writing led to new educational institutions. The creation and copying of manuscripts – usually in soft clay, wood, bamboo, papyrus or parchment – were costly. Thus the second revolution in information technology occurred when Ts'ai Lun invented paper at the turn of the first century. It was one of the key factors for the rapid rise of the Chinese civilization during the next millennium and only became known much later in the West.

The third revolution occurred around 1440 when Gutenberg invented the printing press. This new information technology dramatically reduced the costs of copying information and was a key enabler for the rise of science (Eisenstein 1983, pp 185-252). Before Gutenberg's invention Europe had been in stagnation for more than a millennium. The knowledge of ancient Greek philosophers had been lost and just rediscovered. The number of universities was small (see Figure 1). Eminent thinkers occupied themselves with theology and philosophy. Although paper had already become the dominant medium for the storage of knowledge, copying of manuscripts had to be performed manually. This was costly and error-prone. Direct access to knowledge was scarce and it was difficult to widely and rapidly disseminate ideas.

Table 1. Revolutions in information technology

Year	Technology
ca. 3500 BC	Writing invented in Mesopotamia
ca. 100 AD	Paper invented by Ts'ai Lun
ca. 1440 AD	Printing press invented by Gutenberg
ca. 1945 AD	Electronic computer

The printing press revolutionized the creation and dissemination of knowledge. Forty years after Gutenberg's invention, printing presses had sprung up everywhere in Europe (Steinberg 1959, p 28). Now the renaissance of ancient thought could be completed: "Within fifty years, nearly every important Greek and Roman work had been printed and distributed all over the learned world" (Van Doren 1991, p.154). The number of university foundings rose sharply (see Figure 1).

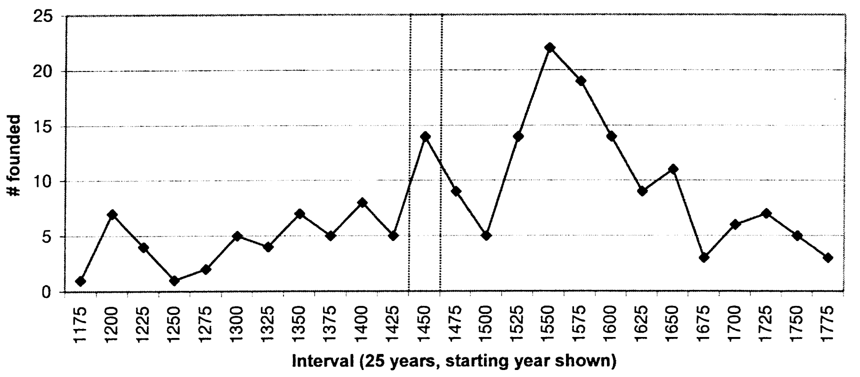

Figure 1. University foundings in Europe 1175-1799 (Data from: Frijhoff 1996, pp 90-94)

The amount of knowledge to which scholars had access exploded and the problems of correct translation and interpretation received much interest. Ideas could be published easily. In this climate, a new era of intellectual discourse began. The explosion of available knowledge enabled the development of new methods, processes and institutions for the creation of a reliable body of knowledge. It favored rational discourse and proof. As a consequence, modern science slowly emerged.

While information technology may not have been the only factor - movable type was known in China and Korea well before Gutenberg's invention (Hart 1992, p 43) - modern science would not have been possible without the printing press.

Currently, the next revolution in information technology is well under way. Like the previous revolutions it will lead to major changes. In contrast to prior revolutions, the causes are not only based on cost reduction, but also on novel capabilities. The introduction of paper reduced the costs of recording information by replacing more costly materials such as parchment, bamboo and papyrus. Similarly, the invention of the printing press reduced the costs and time needed for reproducing manuscripts. While computer technology further reduces the cost of storage and reproduction, it also provides entirely new capabilities: Information can be transmitted instantly over long distances; knowledge which is stored electronically can be modified even after the time of publishing. In addition, it is possible to interact with knowledge in the form of software and knowledge can be presented dynamically in different ways (e.g. as chart or figure) depending on the needs of the reader. Moreover, sound and moving images can be captured. It follows that the present revolution might lead to even more dramatic change than the previous two revolutions.

1.1.2. Potential changes to scientific knowledge

While information technology introduces many new capabilities, it is not clear to what degree scientific knowledge will be affected. A brief analysis may show some of the areas in which the characteristics of scientific knowledge may changes as a consequence of information technology.

Information technology may change the unit used for storing and manipulating scientific knowledge: In the paper environment, the main unit of academic knowledge is the article. An article is a collection of many thoughts and insights and typically contains between two thousand and eight thousand words. For economic reasons, paper journals cannot accept contributions that consist of just a few words or sentences. While it would be economically feasible to distribute such knowledge, the search and integration costs on the side of the reader are prohibitive. On the other hand, large book-sized contributions lead to significant publishing costs. As a consequence, paper journals are forced to limit the size of contributions and thereby may prevent relevant knowledge from being published. For example, evidence in the form of data is routinely suppressed from paper publications. In an electronic environment the costs of publication are nearly zero (see section 2.4.4.5.). This eliminates the upper boundary on article size. In addition, the ability to link information and provide alternative indexes and interfaces can substantially reduce the search costs for the reader. This reduces the limitations on the lower boundary of article size: the individual thought may become the smallest unit of publishable knowledge. In an online environment, therefore, the scientific body of knowledge may consist of a sea of individual scientific thoughts rather than a collection of articles.

In addition, information technology may change the form which scientific knowledge can take. Besides the addition of sound and moving images, scientific knowledge may become active in the form of software. Theories, models and artificial worlds can become alive on the screen. Early examples are already available: An online article by Dolin et al. (1988), for example, includes a section where readers can execute a text classification algorithm and apply it for their own purposes. Similarly a theory of financial decision-making might be formulated in software. The reader might no longer need to understand the theory in detail but could interact with it to apply it directly to his special case. Interactive scientific knowledge may change the dialogue between researcher and reader. It may eliminate the gap between theory and practice which has shielded research from application. At the same time, it may significantly raise the bar for the quality of a theory.

Furthermore, the electronic medium allows the utilization of algorithms for integrating knowledge into a larger body of knowledge and for checking the consistency between different contributions. Overlap with other contributions can be detected in many cases. It may be possible to generate questions aimed at improving structure and content of a contribution which at the same time can be used to check relationships with other contributions.

These changes to knowledge require the institutionalization of new approaches for the creation and dissemination of knowledge. As a consequence, the scientific community may need to adapt the processes of knowledge creation and dissemination to the new technology (Watson1994). At issue is not just information technology as a new medium for the storage of knowledge. Although the ability to modify stored knowledge sets it apart from traditional storage media, the bigger challenge is how to leverage the ability of information processing to improve the body of scientific knowledge. New methods and indexing systems must be accepted by the research community to be effective. While knowledge has traditionally been "pushed" at the audience, in the electronic environment readers can now pull together the knowledge they are interested in from various sites. The processing capability of computers can be used in many other ways to improve not only the retrieval but also the creation, integration and evaluation of knowledge.

This analysis has shown, that significant potential exists for reinventing the scientific communication system. In particular, online knowledge infrastructures may broaden their support from the dissemination of knowledge to the other phases of the research process, including creation, synthesis and evaluation of knowledge.

1.2. Analytical focus

This study is motivated by the transformation of the scientific publishing system which results from the emergence of computer technology. Unfortunately the term ‘scientific publishing system’ may not be well suited to describe the end product of this evolution. Information technology may change the focus away from publishing and dissemination towards discursive knowledge creation and synthesis. To fill this terminological void the term ‘online scientific knowledge infrastructure’ will be utilized in the following. The concept of ‘infrastructure’ is useful because it extends beyond technology. Technology forms the basis of an infrastructure but is never sufficient by itself. An infrastructure requires processes and links with the conventions of practice. It cannot be imposed but grows from the actions of many. An infrastructure must be adopted widely. For the individual, an infrastructure becomes part of the environment of which he becomes aware only when it fails (Star & Ruhleder 1996, p 113).

The current scientific publishing system has many characteristics of an infrastructure: it is widely accepted and most scientific activities are deeply embedded within it. For many researchers the publishing system has become so 'natural' that it is difficult to envision alternatives.

The object of this research thus is a future state of the scientific communication system where the potentials of information technology for the creation and dissemination of scientific knowledge are more fully exploited. While significant progress has been made, this state has not yet been reached. It requires both careful adaptation of information technology towards the needs of scientific processes and the fine-tuning of scientific processes towards the capabilities and requirements of information technology. Ideally, information technology would move into the background and assume the role of a substrate on which research and scientific communication can be performed. It would be invisible to the researcher who could exploit it in many different ways.

As a consequence, the term 'online scientific knowledge infrastructure' will be defined as:

> a socio-technical information system aimed at the creation, organization, storage and application of scientific knowledge. It is not only a technical system consisting of hardware, software, databases etc. It is also on organizational system which consists of stakeholders, organizational structures and processes.

The term 'knowledge infrastructure' needs to be distinguished from similar terms such as 'knowledge base' and 'knowledge repository'. Although 'knowledge base'/'knowledge repository' are occasionally used in the context of organization science to refer to the collective body of knowledge which an organization is able to bring to bear on its operations (Bierly & Chakrabarti 1996) they are predominantly used in a more technical sense. In product development, the term 'knowledge base' refers to computer systems which help engineers to reuse solutions for recurring problems (Borst, Akkermans & Top 1997). In AI it refers to a component of an expert system (or any system that imitates the reasoning behavior of humans). This component supplies 'knowledge' needed for solving a problem to a reasoning engine (Guida & Tasso 1994). A knowledge base or knowledge repository, thus, may be a component of a knowledge infrastructure.

Recently some authors have introduced the term 'knowledge network' (Director & King, 2001). This term can be regarded as synonym to knowledge infrastructure.

1.3. Objectives

Many scientific institutions have begun building online systems dedicated to scientific knowledge. They require a better understanding of the characteristics of online knowledge infrastructures and the choices which need to be made during their design. This research aims to provide a conceptual and empirical basis for the many efforts which are currently under way to use information technology to support the creation, integration, and dissemination of knowledge within science. The ultimate goal of this research is to increase the effectiveness of the scientific enterprise by providing a better understanding of the impact of information technology on the scientific communication system.

1.4. Approach

The second chapter examines characteristics of scientific knowledge infrastructures. This requires both a theoretical analysis of the role which information technology plays in scientific communication as well as an empirical analysis of current efforts which leverage information technology for improving the body of scientific knowledge. Therefore the chapter begins with a theoretical analysis of the transition from traditional paper-based journals towards online infrastructures. The impact of information technology on the main characteristics of paper-based journals and on the associated processes is evaluated systematically. The analysis will show that information technology has the potential of fundamentally changing the scientific communication system. In the second part of this chapter, an analysis of current online infrastructures follows. Internet sites which are dedicated to scientific knowledge are examined and their characteristics are identified. To complete the identification of characteristics, visions of future knowledge infrastructures that are different from the results of theoretical analysis and from currently implemented infrastructures are briefly reviewed. Next the sets of characteristics identified in the first three sections of chapter 2 are examined in detail. Groups of similar characteristics are identified. The main groups are 'mode of interaction', the 'structure of knowledge', 'presentation', 'governance' and 'technology'. Next the alternatives, which are available for the implementation of each characteristic, are evaluated. The result is a comprehensive matrix of characteristics which can be used for categorizing knowledge infrastructures and which can be consulted for design decisions.

The third chapter addresses a core problem which all scientific knowledge infrastructures face: Finding an adequate structure for representing scientific knowledge within the infrastructure. The problem arises independently of whether a knowledge infrastructure is to be based on paper or information technology. It is complicated by the lack of agreement among scientists as to what constitutes scientific knowledge. Therefore, the purpose of the third chapter is to develop a conceptual model of scientific knowledge which can be used as the basis for the implementation of knowledge infrastructures. First the objectives of developing a structure of scientific knowledge are examined. Then the concept of scientific knowledge is examined from a theoretical and philosophical perspective. This includes an examination of different interpretations of scientific knowledge and of the impact of these conceptualizations on scientific paradigms. Subsequently an object-oriented model of scientific knowledge in the form of a class diagram is developed. Different categories of scientific knowledge are identified by examining the categories which are used by eminent philosophers of science such as Popper, Nagel, Dubin and Bunge and in more practice-oriented text books on research methodologies. The relationship between these categories and their attributes are examined subsequently. This yields a detailed model of scientific knowledge that can be used as the foundation for implementing a knowledge infrastructure.

The last chapter discusses the implications of this research. To show the feasibility of knowledge infrastructures which significantly depart from the traditional model, a system which has been developed by the author and is available on the internet will also be discussed.

http://www.springer.com/978-3-540-01250-4

From Publishing to Knowledge Networks
Reinventing Online Knowledge Infrastructures
Hars, A.
2003, XVII, 211 p., Hardcover
ISBN: 978-3-540-01250-4