Sharing and Exchanging Knowledge in Communities: The CYCLADES Open Collaborative Virtual Archive Environment

Tom Gross
GMD-FIT, GMD — German National Research Center for Information Technology
(tom.gross@gmd.de)

Abstract. Since the 1940s several great analyses of the state-of-the-art and predictions and visions of further research and development in information and communication technology have been published. From early on both the individual knowledge worker and on-line communities have been focal points of their reflections. Today, a considerable share of the early visions has already been put into practice and we are increasingly facing ubiquitous availability and pervasive use of information and communication technology. Nevertheless, information technology on the one hand and communication technology on the other are often developed and used in isolation from each other. In this paper I argue for a stronger convergence of support for individual knowledge workers and of online communities of knowledge workers. I will discuss some basic notions, revisit related studies, derive system requirements, and exemplify the ideas by presenting the CYCLADES open collaborative virtual archive environment.

1 Introduction

Since the 1940s several authors have published great analyses of the state-of-the-art of their respective time as well as predictions and visions of further research and development in information and communication technology. From early on both the individual knowledge worker and on-line communities have been focal points of these reflections. For instance, Vannevar Bush [1945] had great and fundamental thoughts on human thinking and adequate support for individual users by introducing associative memories. Douglas Engelbart [1963] developed a very early text processing system that allowed users to start video conferences while working with their text. Joseph Licklider [1960; 1968] invented the metaphor of the computer as a communication device and aimed at supporting on-line communities of interest.

Today, we increasingly face ubiquitous availability and pervasive use of information and communication technology. A considerable share of the early visions has already been put into practice. Information technology in a broader sense of its meaning allows users to

flexibly exchange and share documents, for instance, via Internet services such as FTP and the WWW, via shared global workspaces such as BSCW [Bentley *et al.* 1997], but also via peer-to-peer systems such as Napster and Gnutella [O'Reilly 2001]. Communication technology such as email, instant messaging, and short messaging via mobile phones has spread tremendously. However, these two classes of systems—information technology on the one hand and communication technology on the other—are often developed and used in isolation from each other. In this paper I argue for improving support of individual knowledge workers and online communities of knowledge workers through a stronger integration of information and communication technology. For this purpose I start with a short clarification of my understanding of some epistemological notions such as data, information, and knowledge. I will summarise related studies analysing the sharing and exchange of knowledge and derive some requirements for systems supporting flexible sharing and exchanging of knowledge in communities. Finally, the CYCLADES open collaborative virtual archive environment is presented to exemplify some of these ideas.

2 Knowledge

Although, I regretfully missed the first workshop on 'Community Knowledge' I want to strongly agree with the view of community knowledge having an 'alpha' compared to the individual members of the community. Although a thorough discussion of epistemology would go beyond the scope of this position paper, it is important to clarify some basic notions of human knowledge. In this paper I will use the notions as follows: data is raw; information is data with meaning; and knowledge is verified information [Gross 1997]. Expertise is seen as the 'embodiment of knowledge and skills within individuals' [McDonald & Ackerman 1998, p. 315]. Whereas data and information can be easily processed by and exchanged via computers, the verification that is inherent in knowledge is typically a human activity, which can only be delegated to computers and mediated through information systems and knowledge management system to some extent. Expertise is *per definitionem* human. Therefore, sharing and exchanging knowledge and expertise cannot be automated totally.

These aspects have sometimes been neglected. For instance, classic information theory took a probabilistic approach (e.g., Shannon and Weaver [1963], Carnap [1962]). Later, several authors contradicted the information theory of Shannon and Weaver and primarily focused on the processing aspect (e.g., Newell and Simon [1972]). In the 1990s the ecological concept has been trying to combine aspects of these two concepts into a more holistic approach (e.g., Devlin [1991]). In some areas of computer science this ecological concept has been adopted. For instance, in parts of the human-computer interaction community it is used in situation theory. Here humans (and to some extent robots) are seen as systems that have to adapt to their environment and therefore constantly have to capture information about their environment and process it (e.g., Maturana and Varela [1984]). An important point in situation theory is that constraints vary in different environments and situations and that therefore, information and knowledge are dependent to the respective setting. Jean Lave calls these situative specifics and understanding under these conditions 'local epistemologies' [1991]. As opposed to computers human experts often easily—sometimes subconsciously—adapt to the respective settings.

3 Sharing and Exchanging Knowledge

With these notions in mind I now start analysing the human activities involved in sharing and exchanging knowledge. For this purpose I will revisit some studies found in literature.

Twidale and associates [1994] report on an informal observation of a group of students using OPAC terminals. They found that students did free queries as solitary activities, but also had chance encounters with other students. Furthermore, they did coordinated search and also joint search. Ackerman [1994] enumerates similar forms of social interaction among information seekers in digital libraries.

In another study Twidale and associates [1997] identified several tactics of information seekers. Information seekers can consult colleagues and ask them for information and references; they can wander around to search for information and hope to meet others coincidentally; they can do brainstorming to generate numerous ideas; or they can bibble—that is, they can look for existing bibliographies to the same or a similar topic (this strategy is often used on the WWW, when users use the bookmark list of other users with similar interests).

This empirical evidence shows that—amongst other consequences—flexible switching between solitary work and information seeking and spontaneous contacts and dialogues with others are a core requirement for flexible sharing and exchange of knowledge. Kuhlthau [1991] developed the notion of zone of intervention as an area, in which this is possible. This zone of intervention is an area in which users are very efficient in gathering the information needed. She defines a zone of intervention as:

> …that area in which an information user can do with advice and assistance what he or she could not do alone. Intervention within this zone enables the user to move along in the information search process. Intervention outside this zone is inefficient and unnecessary; experienced by users as intrusive, on the one side and overwhelming, on the other.

The results of these studies have clear implications for the design of systems that aim at supporting the flexible sharing and exchanging knowledge. First, technology has to extend the reach of the single knowledge worker; so, she can easily and flexibly contact other persons and share and exchange knowledge when needed. Secondly, technology needs to empower communities of knowledge workers to establish a common knowledge base over time—often called community memory. Marshall and associates [1994] argue that 'when people work together—whether in designing a product, or creating training materials from video-based documentation, or writing a coherent analysis of a complex situation in the world—they require, and put effort into constructing and maintaining, shared understandings of what they are doing'. They define a community memory as an 'open-ended set of shared interpretations and understandings developed and maintained by the group'. In order to create, maintain, and increase community memory, systems have to support the acquisition and continual updates of the contents and the structure of the community memory and the identification of the relevancy of material found. As a result the community memory reflects the evolution of shared understanding.

4 The CYCLADES Open Collaborative Virtual Archive Environment

The EU-project CYCLADES aims at developing an open collaborative e-print archive environment to support online communities of scholars [Gross 2001]. Electronic pre-print (e-print) archives are important vehicles for the dissemination of preliminary results and non-peer reviewed grey literature. Most of them focus on the dissemination of information within disciplinary or institutional communities. Scientific research, however, increasingly tends to be interdisciplinary. There is thus a growing need for easy retrieval of information from diverse sources, and for communication and collaboration across traditional community boundaries. CYCLADES addresses these issues and supports the transition of e-print systems into genuine building blocks of a transformed scholarly communication model by developing a set of leading edge technologies.

4.1 Functionality

The CYCLADES prototype supports scholars both individually and as members of online communities of scholars when interacting with very large virtual e-print archives. It supports *individual scholars* in efficiently and effectively retrieving relevant information from a very large, distributed, heterogeneous, multi-disciplinary and highly dynamic archives environment. It gives feedback on the degree of relevance of the retrieved information. It informs users about new incoming information in the archive environment based on their interests and timely preferences. It supports *communities of scholars* with a platform for sharing and exchanging knowledge. It allows users to share documents and annotate them in shared web-based workspaces. It disseminates relevant information to community members in the form of recommendations that are based on the analysis of user behaviour and collective profiles. It provides functionality for user ratings of shared documents. On a whole, it allows community members to learn from, contribute to, and collectively build upon, the community's knowledge.

4.2 Architecture

The CYCLADES environment consists of two components: the archive environment and the service environment. The *archive environment* is composed of a large number of heterogeneous, multidisciplinary archives and supports interoperability between them. The *service environment* is composed of a number of autonomous but interoperable services which support a scholar as well as a community of scholars when carrying out their work. Each of these services provides unique functionality. A service may work independently, or can communicate and collaborate with the other services in order to provide new, value-added services. The development of the service environment is the main objective of the CYCLADES project. Figure 1 shows a functional view on CYCLADES prototype with the single scholar and the communities on top and the service environment underneath and the archive environment at the bottom.

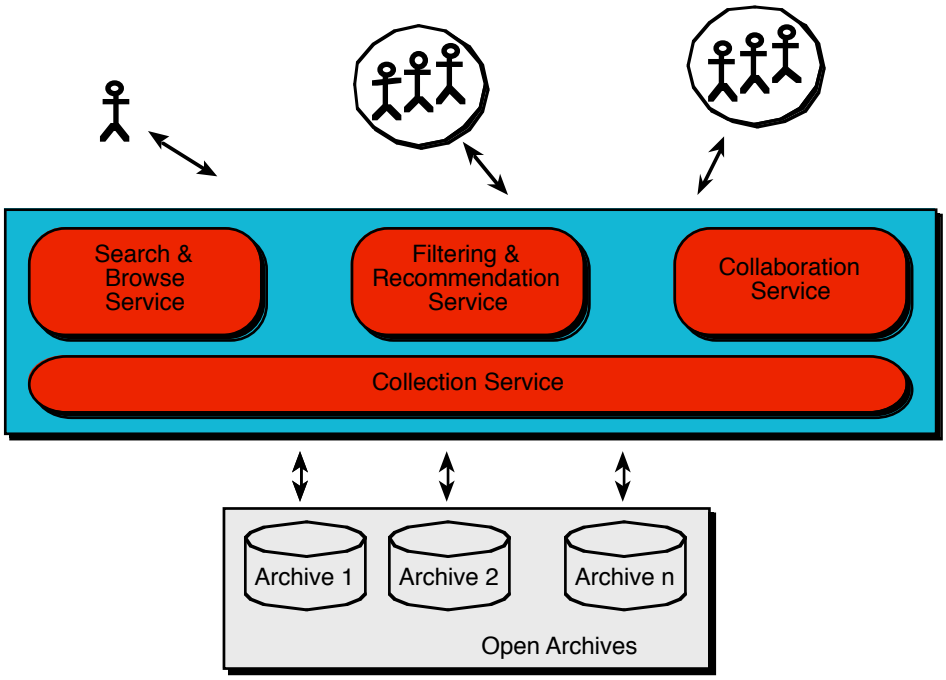

Figure 1. A functional view on the CYCLADES prototype.

The partners of this IST-project of the EU are the European Research Consortium for Informatics and Mathematics ERCIM, the Italian National Research Council CNR, the Foundation for Research and Technology Hellas FORTH, the German National Research Center for Information Technology GMD, and the University of Dortmund UNIDO. GMD-FIT contributes knowledge and expertise concerning the design and implementation of Internet-based shared workspaces as well as rating and recommendation systems.

5 Conclusions

In this workshop I would like to discuss the notions of knowledge with respect to the individual knowledge worker and knowledge with respect to communities of knowledge workers. I am particularly interested in the implications of this perspective on the analysis and design of cooperative systems. Finally, I would like to learn about existing systems supporting the flexible sharing and exchange of community knowledge.

6 Biographical Information

Tom Gross holds a diploma and a doctorate degree in Applied Computer Science from the Johannes Kepler University Linz, Austria. Since 1999 he is a senior researcher in the Computer-Supported Cooperative Work research group Institute for Applied Information Technology FIT at the German National Research Center for Information Technology. He is coordinating the activities of project CYCLADES at FIT. His research interests include computer-supported cooperative work, human-computer interaction, and global information systems.

7 Acknowledgements

The research presented here was carried out and financed by the IST-2000-25456 project CYCLADES, partly funded by the EC. I would like to thank all my colleagues from the CYCLADES team.

8 References

Ackerman, M.S. Providing Social Interaction in the Digital Library. In Proceedings of the Digital Libraries Workshop - DL'94 (May 19-20, Newark, NJ). Springer-Verlag, Heidelberg, 1994.

Bentley, R., Appelt, W., Busbach, U., Hinrichs, E., Kerr, D., Sikkel, S., Trevor, J. and Woetzel, G. Basic Support for Cooperative Work on the World Wide Web. Human-Computer Studies 46, 6 (June 1997). pp. 827-846.

Bush, V. As we may think. The Atlantic Monthly 176, 1 (July 1945). pp. 101-108.

Carnap, R. Logical Foundations of Probability. University of Chicago Press, London, UK, 1962.

Devlin, K. Logic and Information. Cambridge University Press, Cambridge, MA, 1991.

Engelbart, D. A Conceptual Framework for the Augmentation of Man's Intellect. In Howerton, P., ed. Vistas in Information Handling. Spartan Books, Washington, DC, 1963. pp. 1-29.

Gross, T. Towards Bridging the Gap Between Humans and Information Systems. In Proceedings of the Fifth International Information Management Talks - IDIMT'97 (Oct. 15-17, Zadov, Czech Republic). Oldenbourg, Vienna, 1997. pp. 147-156.

Gross, T. CYCLADES: An Open Collaborative Virtual Archive Environment. http://fit.gmd.de/projects/cyclades/, 2001. (Accessed 7/8/2001).

Kuhlthau, C.C. Inside the Search Process: Information Seeking from the User's Perspective. Journal of the American Society for Information Science 42, 5 (May 1991). pp. 361-371.

Lave, J. Situating Learning in Communities of Practice. In Resnick, L.B., Levine, J.M. and Teasley, S.D., eds. Perspectives on Socially Shared Cognition. American Psychological Association, Washington, DC, 1991. pp. 63-82.

Licklider, J.C.R. Man-Computer Symbiosis. IRE (now IEEE) Transactions on Human Factors in Electronics HFE-1 (Mar. 1960). pp. 4-11.

Licklider, J.C.R. The Computer as a Communication Device. Science and Technology, (Apr. 1968).

Marshall, C., Shipman, F.M. and McCall, R.J. Putting Digital Libraries to Work: Issues from Experience with Community Memories. In Proceedings of the Digital Libraries Workshop - DL'94 (May 19-20, Newark, NJ). Springer-Verlag, Heidelberg, 1994.

Maturana, H.R. and Varela, F.J. El arbol del concocimiento (Der Baum der Erkenntnis: Die Biologischen Wurzeln des Menschlichen Erkennens; Originally Spanisch, German translation). Goldmann Verlag, Bern, Switzerland, 1984.

McDonald, D.W. and Ackerman, M.S. Just Talk to Me: A Field Study of Expertise Location. In Proceedings of the ACM 1998 Conference on Computer-Supported Cooperative Work - CSCW'98 (Nov. 14-18, Seattle, WA). ACM, N.Y., 1998. pp. 315-324.

Newell, A. and Simon, H.A. Human Information Processing. Prentice-Hall, Engleweed Cliffs, NJ, 1972.

O'Reilly. OpenP2P.com. O'Reilly & Associates, Inc., http://www.openp2p.com/, 2001. (Accessed 7/8/2001).

Shannon, C.E. and Weaver, W. The Mathematical Theory of Communication. University of Illinois Press, Urbana, IL, 1963.

Twidale, M.B., Mariani, J., Rodden, T. and Sawyer, P. Awareness as a Mechanism for Supporting the Learning of Database Browsing Techniques. Computing Department, Lancaster University, http://www.comp.lancs.ac.uk/computing/research/aai-aied/information/ariadne/docs/original.html, 1994. (Accessed 11/6/1999).

Twidale, M.B., Nichols, D.M. and Paice, C.D. Browsing is a Collaborative Process. Information Processing and Management 33, 6 (1997). pp. 761-783.