



Good Read: Data at Work: Supporting Sharing in Science and Engineering

- Mason Kortz (CCE/PAL)

Birnholtz, Jeremy and Matthew Bietz, Data at Work: Supporting Sharing in Science and Engineering. Conference on Supporting Group Work. (2003) 339-348

As information managers, we are all acutely aware that there are social barriers to data sharing. In this paper, authors Birnholtz and Bietz discuss the origins of some of these social barriers, and provide suggestions on how those barriers can be addressed from a Computer Supported Cooperative Work (CSCW) system design standpoint. To do so, they first explore the various ways in which data are valued in scientific communities. From there, they discuss the ways in which this value may be enhanced or compromised by data sharing. They end with a set of considerations for developing collaborative data systems, and a brief summary of data sharing issues in need a further detailed research.

The methodology for the research leading to this paper is a set of ethnographic studies across three disciplines – HIV/AIDS research, earthquake engineering, and space physics. This interview process allows the authors to identify both common and divergent themes in data sharing. It also means that the viewpoints on the value of data presented as described by researchers themselves. Two broad roles are identified for data: as scientific evidence and as a social construct in the community. The latter role is explored more extensively, and is further broken into sub-roles. This discussion includes the value of data in defining communities of practice, in establishing relationships within communities, and as an indicator of status. The differences in how data are valued in various communities are pointed out and drawn back to qualities of those communities including task uncertainty, feasibility of single-lab science, and academic tradition.

Having established the ways in which data are valued by researchers, the authors address the impact on data sharing practices. Data are described as objects that have the potential to generate various revenues – status, publication, funding, etc. By sharing data, scientists have the potential to gain revenues by entering collaborations that exceed the scope of what can be achieved in one lab. However, this exposure also presents the risk of data being misused, of mistakes in the data being made public, and in the data provider getting 'scooped' on a publication. Also discussed is the need for the context of the data to be shared. While metadata is acknowledged as an important part of context, it is also recognized that metadata is rarely complete and never easy to generate.

The authors close with some brief recommendations on the design of CSCW systems as well as future research into data sharing practices. Design suggestions include building support community-specific social constructs into collaborative systems, recognizing the multiple roles of data and supporting them appropriately, and not relying solely on metadata to enable sharing, but also to support sharing of broader contextual information. Further studies would include research on the role of data abstractions – specifically on how they can maximize the benefits of data exposure while minimizing risks – and on the further development of metadata or other contextual information.

While the themes present in this paper are not unknown to our information management community, it is helpful to see them presented in a broader framework. For participants who are new to the practice of information management or who do not have a background in scientific research, this paper makes very clear the value of being aware of data as more than simply a tool for publication and as part of a knowledge-making process. Understanding the values perspectives of a community is essential to the design of systems that will support that community, and thus essential to the role of the information manager.