In their article "Data at Work: Supporting Sharing in Science and Engineering," Birnholtz and Bietz (2003) argued that the information science field needs "a better understanding of the use of data in practice" (p. 339). Collaborating and sharing data within and between groups of scientists is something that needs to be encouraged, they said, but "sharing data is not easy" (p. 340) for at least three reasons: (a) a lack of willingness to share and trust of others; (b) problems with finding data that has been shared; and (c) difficulties with interpreting and using data shared by others. While their article did feature some bias and was not always completely clear, their study and its conclusions is important and significant for the field of information science, particularly for the design of information systems and research into social information behavior.

Summary

Birnholtz and Bietz (2003) used observation, interviews, and surveys to study the data practices of earthquake engineers, HIV/AIDS researchers, and space physicists. They found that there were two dimensions of data use which differed between—and sometimes within—these disciplines. First, data could confirm existing results or be seen as "news" of something unexpected that required further exploration (p. 341). Second, data could be focused on an "event," as in a cross-sectional study, or exist as a "relatively constant stream" relating to longitudinal work (p. 341).

The authors also discussed the relationship between data and scientific communities, identifying four categories of possible data uses (with an eye to "revenue" from such uses). They defined these (Birnholtz & Bietz, 2003, p. 345) as

> (1) "a scientist's data set is her [or his] castle" - the researcher has the ability to use the data and wants to use it for a particular problem or question;
> (2) "with a little help from my friends" - the researcher wants to use the data, but needs to collaborate with others in order to be successful;
> (3) "one scientist's junk is another one's treasure" - the researcher has no interest in using the data for a particular problem or question, but other researchers do have interest in the data and that problem; and
> (4) "d'oh!" - the researcher hasn't thought of using the data in a way that is relevant to a problem or question he/she is interested in.

Birnholtz and Bietz argued that researchers would be less willing to share data unless there were high incentives for data uses falling within categories 2 and 3. Sharing of and communication around abstractions of data sets, they felt, would produce sufficient incentive for those categories without scientists having to give away all their data and run the risk of someone else "scooping" their results (category 1) or embarrassing them (category 4). They concluded by recommending that efforts to support "social interaction [and communication] around data abstractions and the data themselves" should be made, that metadata should be augmented through "the sharing of supplementary materials," and that consideration of the "social and scientific roles of data" was necessary in future research.

Critique

Birnholtz and Bietz's arguments were generally well-supported by the examples they gave, although their article and studies were clearly biased towards the natural sciences. Their study of scholars in engineering, physics, and biomedicine does not necessarily apply to how social scientists might share and use data. While the article was generally coherent, a few of the authors' points were a little difficult to follow and could have been expressed more clearly. In particular, the examples given from AIDS research and earthquake engineering of how "data enable inbound trajectories" were claimed to be different by Birnholtz and Bietz, exposing what they called a "critical" distinction (p. 344). However, the difference—assuming there actually was one—was subtle and easy to miss. They also promised they would "identify three aspects of the way data are used that [would] impact how data get shared" (p. 341), but then only discussed two dimensions. Finally, they did discuss the limitations of their study, but unfortunately only very briefly.

Despite these minor flaws, the overall conclusions of Birnholtz and Bietz are important to information science as a field, and particularly relevant to the areas of information systems design and information behavior. Research such as theirs that helps support the sharing of data and information—as well as collaborative research based on such data and information—is vital both to move the collaborative work of scholars forward and to help design better information systems that serve scholars' information needs, both individually and in social context. Such research should aim to better understand the individual and collaborative information behaviors of scholars from all disciplines—natural science, social science, humanities, and interdisciplinary fields—with particular emphasis on how and why these scholars share—or do not share—data, information, and knowledge with others and on how such sharing can be encouraged and better supported.

Reference

Birnholtz, J., & Bietz, M. (2003). Data at work: Supporting sharing in science and engineering. In M. Pendergast (Ed.), *GROUP '03: Proceedings of the 2003 international ACM SIGGROUP Conference on Supporting Group Work* (pp. 339-348). New York, NY: ACM.