

The International Journal of Digital Curation

Issue 1, Volume 6 | 2011

Beyond the Data Deluge: A Research Agenda for Large-Scale Data Sharing and Reuse

Ixchel M. Faniel and Ann Zimmerman,
School of Information,
University of Michigan

Abstract

There is almost universal agreement that scientific data should be shared for use beyond the purposes for which they were initially collected. Access to data enables system-level science, expands the instruments and products of research to new communities, and advances solutions to complex human problems. While demands for data are not new, the vision of open access to data is increasingly ambitious. The aim is to make data accessible and usable to anyone, anytime, anywhere, and for any purpose. Until recently, scholarly investigations related to data sharing and reuse were sparse. They have become more common as technology and instrumentation have advanced, policies that mandate sharing have been implemented, and research has become more interdisciplinary. Each of these factors has contributed to what is commonly referred to as the “data deluge”. Most discussions about increases in the scale of sharing and reuse have focused on growing amounts of data. There are other issues related to open access to data that also concern scale which have not been as widely discussed: broader participation in data sharing and reuse, increases in the number and types of intermediaries, and more digital data products. The purpose of this paper is to develop a research agenda for scientific data sharing and reuse that considers these three areas.¹

¹ This paper is based on the paper given by the authors at the 6th International Digital Curation Conference, December 2010; received December 2010, published March 2011.

The *International Journal of Digital Curation* is an international journal committed to scholarly excellence and dedicated to the advancement of digital curation across a wide range of sectors. ISSN: 1746-8256 The IJDC is published by UKOLN at the University of Bath and is a publication of the Digital Curation Centre.





Introduction

There is almost universal agreement that scientific data should be shared for use beyond the purposes for which they were initially collected. Access to data enables system-level science, expands the instruments and products of research to new communities, and advances solutions to complex human problems (e.g., Hey & Trefethen, [2003](#); Hey, Tansley, & Tolle, [2009](#)). While demands for data are not entirely new, the vision of open access to data is increasingly ambitious. The aim is to make data accessible and usable to anyone, anytime, anywhere, and for any purpose.

Until recently, scholarly investigations related to data sharing and reuse were sparse. They have become more common as technology and instrumentation have advanced, policies that mandate sharing have been implemented, and research has become more interdisciplinary. Each of these factors has contributed to what is commonly referred to as the “data deluge” (Hey & Trefethen, [2003](#)). Most discussions about increases in the scale of sharing and reuse have focused on growing amounts of data. We identified three other issues related to open access to data that also concern scale that have not been widely discussed:

- broader participation in data sharing and reuse;
- increases in the number and types of intermediaries;
- more digital data products.

These three areas form the basis of a research agenda we developed for scientific data sharing and reuse. In this paper, we describe each issue in more detail, discuss findings from prior research and identify research questions that need to be addressed. The aim of the agenda is provide a roadmap to help build a cumulative body of knowledge that furthers basic understanding and informs practice. We do not prescribe specific conceptual frameworks or approaches because we believe multiple perspectives and methods are needed to address the questions we pose.

Broader Participation in Data Sharing and Reuse

Broader participation in data sharing and reuse is an important issue in considerations of scale and can be viewed from three perspectives. First, data sharing and reuse are becoming important in domains in which they were previously uncommon. This situation provides opportunities to understand the factors that drive sharing and reuse among members of the same science community, and the influence these activities exert on culture, practice and communication in fields where they are new. Second, the focus on interdisciplinary research, along with non-scientist participation and interest in the research process, means that data are being used by individuals who are outside of the community in which they were generated. These new contexts of reuse raise questions about how individuals from different cultures and with varied knowledge and expertise find, understand, and reuse data. Finally, in addition to their role as data reusers, non-scientists are increasingly collecting, sharing, and analyzing data that may be used to study scientific questions. This situation has implications for broader participation in science and for understanding how scientists trust and reuse data collected by non-scientists. In the sections that follow we discuss prior work in these areas as well as open questions driven by broader participation in sharing and reuse.



Data Sharing and Reuse within Science Communities

Much of the prior research has focused on scientists' motivations to share data with others in their own community. These studies show that fields in which data sharing is common are characterized by a mixture of technical capabilities, such as free and easy software for data transfer, management and analysis; socially influenced demands and incentives; and scientifically motivated needs, especially the questions that scientists want to answer (e.g., Birnholtz & Bietz, [2003](#); Griffiths, [2008](#)). The latter factor is particularly important. For example, research in physical oceanography is conducted using large research vessels carrying expensive data collection equipment. Data are gathered from remote locations, which requires coordination across long distances. Data sharing and reuse are necessary to conduct physical oceanographic research since no individual and few institutions can afford to carry it out on their own (Hesse, Sproull, Kiesler, & Walsh, [1993](#)). For other disciplines, it is only recently that the science being conducted requires data from others within their disciplines. In ecology, for instance, changes to sharing and reuse patterns are being driven by the collection of new types of data, by the online availability of large volumes of data of interest to ecologists, by technologies that make it easier to manage and integrate disparate data, and by slowly changing views about the value of secondary analysis to address important ecological questions (Borgman, Wallis, & Enyedy, [2007](#); Zimmerman, [2008](#)). Further studies, exemplified by the questions below, are needed to understand the degree to which prior findings are applicable to various disciplines and to make comparisons across disciplines.

1. *How do the factors that motivate data sharing and reuse differ across science communities, and what contributes to the differences?*
2. *What precipitates the cultural changes necessary for science communities to engage in large scale data sharing and reuse?*

The implicit assumption in much of the literature is that making data more widely available will ensure reuse. However, the few studies that have been conducted show that data reuse is difficult even among scientists from the same community. The major challenge to reuse is that data are embedded in a local context, which makes it difficult for reusers to understand and trust the data (e.g., Berg & Goorman, [1999](#); Cragin & Shankar, [2006](#); Jirotko et al., [2005](#); Zimmerman, [2008](#)). There has been some research into how the local context is communicated to reusers, but no definitive conclusions have been drawn. Some studies have found that the documentation scientists produce for themselves can be of limited use to others (Birnholtz & Bietz, [2003](#); Shankar, [2007](#); Zimmerman, [2008](#)). Other investigations have shown scientists' documentation to be quite useful to reusers (Carlson & Anderson, [2007](#); Faniel & Jacobsen, [2010](#)). Still other research has found that social exchange with the data producer is an important part of data reuse (Collins, [1992](#)). However, social exchange is difficult to accomplish on a large scale, in part because it not always possible or desirable for data producers to communicate with data reusers. Furthermore, technology has not yet been successful in bridging this gap. Anecdotal evidence suggests scientific workflows and social media might be useful for documentation and community curation (e.g., De Roure, Goble, & Bhagat, [2008](#)), but very little research has been done to measure their effectiveness for the data producer or reuser. These challenges suggest the following research questions:

- *What other types of social interaction beyond that with the data producer can facilitate data reuse (e.g., colleagues, third party experts)?*
- *How might technology be employed to capture and communicate the local context for reusers, while reducing the burden for data producers?*
- *How can social exchange and documentation be combined to support data sharing and reuse on a large scale?*

Data Sharing and Reuse across Different Communities

The challenges related to data sharing and reuse within a science community become more daunting as these activities open to scientists and non-scientists outside of the community in which the data were generated. Each type of reuser has particular knowledge, skills and practices, as well as different purposes for reusing data, not all of which are research oriented. For instance, educators, practitioners, policymakers and the general public may want to use scientific research for pedagogy, product and service innovations, policy formulation, or hobbies. Below, we discuss issues related to sharing and reuse by scientists who are outside the domain in which the data were produced and by non-scientists.

Scientists are being encouraged to reuse data from multiple domains because interdisciplinary research is believed to be an important part of addressing many of today's complex problems. Interdisciplinary research occurs when knowledge, experience, technology or expertise is transferred via borrowing, collaboration or boundary crossing (Pierce, [1999](#)). The need for interdisciplinary studies is often raised in the context of grand challenge research which seeks to answer pressing questions that impact society, and have the potential to yield major results and practical benefit if addressed (National Research Council, [2010](#)). For example, a national database of mammogram images may be useful to epidemiologists investigating factors that contribute to breast cancer (Jirotko et al., [2005](#)). In addition to the challenges discussed in the previous section, there is little understanding about what needs to be done to facilitate data reuse in such new contexts. Based on what is known about interdisciplinary research, we expect that it will be difficult for reusers to acquire the technical, tacit and theoretical knowledge required to understand and reuse data collected from other fields.

The same factors that make it hard for scientists to reuse data collected by those from a different community also make it difficult for data producers to share data. Even when they are motivated to share, it is difficult for data producers to provide documentation or otherwise communicate with others outside of their community. Members of a discipline share common terminology and methods as well as their own publication channels for disseminating research (Klein, [1996](#)). Scientists who conduct interdisciplinary research face a number of challenges because the disparities between disciplines make it difficult to communicate information across them (Palmer, [1996](#); Pierce, [1999](#)). These differences include the expectations of those involved in the peer review process, the models or paradigms on which research is based, and the distinct stylistic and presentational features that exist in each field (Pierce, [1999](#)). These disparities also lead to concerns by data producers that their data will be misused (Van House, [2002](#); Van House, Butler, & Schiff, [1998](#)).



All of the issues scientists face when reusing data for interdisciplinary research are magnified when non-scientists attempt to reuse science data. Non-scientists have different needs, goals, skills and knowledge. For example, scientists may rely on factors internal to the scientific enterprise (e.g., methodological rigor) when working with data, whereas non-scientists may depend on ones external to the scientific process, such as the extent to which scientific explanations match their experience (Weeks & Packard, 1997). Moreover, it is not clear that non-scientists are interested in reusing the data per se. Instead, they may be interested in reusing the various products of scientific research (e.g., interpretations, discussion of practical implications), or they may benefit from new products that are developed to match their needs (e.g., synthesis documents, guidelines for application).

The last issue we take up in regard to broader participation in sharing and reuse is the production of data by non-scientists. For example, data collected by lay people in astronomy, entomology, botany, cancer research or other fields are potentially valuable to scientists and are sometimes used by them (e.g., Luther, et al., 2009). Questions remain, though, about how scientists come to trust data gathered by non-scientists, even when processes are in place to support sharing across different communities. For instance, when non-scientists began contributing observational data to the California Digital Library, changes were made to procedures for sharing data. Specifically, data producers were asked to report their level of expertise, experience and confidence in making observations, and to submit their observations to third party expert review. However, these processes did not resolve all the concerns of scientist users (Van House, 2003). In other cases, scientists have found ways to confidently reuse non-scientists' data. Some ecologists, for example, have used data on the flowering time of plants and bird migration collected by amateur naturalists to study climate change (Whitfield, 2001).

The sharing and reuse of data across different communities raises a number of questions for future research, including:

- 1. How do data sharing practices within a science community change as non-members participate as data producers? How do non-members become viewed as legitimate participants in data sharing activities?*
- 2. When are non-scientists in need of scientific research, what are they interested in reusing, and how do their reuse practices differ from members of the science community?*
- 3. How do the reuse practices of people who are not members of a science community vary across user type and reuse purpose?*
- 4. What factors influence the degree to which an integrated function for data sharing and reuse among members and non-members of a science community can be offered?*



Increases in the Number and Types of Intermediaries

As we move from small to large scale data sharing, where data are managed and maintained for broad access, we also are seeing an increase in the number and type of intermediaries. Intermediaries, in the form of organizations and the people who work for them, prepare data for reuse by eliciting, organizing, storing, packaging and/or preserving data, and by performing various roles in dissemination and facilitation (Markus, [2001](#)). Three intermediaries that currently exist are data archives, institutional repositories (IRs) and virtual organizations (VOs). Below we discuss the strengths and capabilities of each with regard to data sharing and reuse.

Until recently, data archives that acquire, manage and preserve data intended for use by a specific domain or by the general science and education community were the primary infrastructure for data sharing (Green & Gutmann, [2007](#); National Science Board, [2005](#)). Data archives have staff and expertise that allow them to offer support throughout the data lifecycle, including the capture, management and preservation of data (Borgman, [2007](#); Green & Gutmann, [2007](#)). They also provide infrastructure for data sharing and reuse that includes documentation, statistical services and standardization. Furthermore, data archives that serve specific science domains have close connections to those communities. The Inter-University Consortium for Political and Social Research and the Arabidopsis Information Resource in the biological sciences are two examples of domain specific archives. For disciplines that lack data archives, IRs and VOs offer two different alternatives.

Libraries are building IRs to house many types of digital intellectual products (e.g., publications, presentations) created by the faculty, research staff and students of their institutions (Crow, [2002](#)). Although librarians have become increasingly interested in working with all kinds of scholarly output, IRs are currently best suited to collect, organize and preserve materials near the end of the research life cycle. Examples of IRs include DeepBlue at the University of Michigan and e-Scholar at Purdue University. The advantages of IRs are that the process to contribute to them is simple and they can serve disciplines without other options for data sharing (Green & Gutmann, [2007](#)). In addition, the scholarly publications and other end products of research that libraries collect are often a starting point for data reusers, which make IRs important intermediaries for data (Zimmerman, [2007](#)).

VOs are different still from IRs and data archives. They have been established by science and engineering communities to take advantage of new technologies to support collaboration and provide access to distributed resources such as instruments, tools and data (Cummings et al., [2008](#)). Examples of VOs include the George E. Brown, Jr. Network for Earthquake Engineering Simulation, the Cancer Biomedical Informatics Grid and the Sloan Digital Sky Survey (Cummings et al., [2008](#)). A major objective of a VO is to develop specialized tools and technologies to not only put in and take out data but also work with the data (e.g., analyze, visualize). Although VOs have expertise in the instruments and methods to produce the data, they may have fewer staff and less expertise managing large-scale data archives.

Some people have suggested that the various intermediaries should consider building partnerships and clarifying roles and responsibilities, given differences in their strengths and the ways they facilitate data sharing and reuse (e.g., Association of

Research Libraries, [2006](#); Green & Gutmann, [2007](#)). With this in mind, the overarching research questions we pose focus on the choices among intermediaries from the perspective of those who contribute and reuse data. In addition, we call for more research that examines factors related to the growth and effectiveness of these intermediaries or new ones that might emerge.

What factors influence data producers' decisions about where to deposit data?


- *What factors influence data reusers' seeking behavior and experiences in data reuse?*
- *What contributes to the success of large scale data management, sharing and reuse? How do these factors differ across intermediaries?*
- *What are the affordances of each intermediary from the perspective of data contributors and reusers?*
- *What organizational, social and technical arrangements are needed to manage dependencies and coordinate offerings across intermediaries?*

Along with the expansion of intermediaries in the sense of institutions and organizations, there are more individuals who have a part to play in the data universe. Large-scale data sharing and reuse requires skills and expertise that span the entire lifecycle of data. In addition, some people are stepping into newly created roles and others are finding they need to redefine existing ones. For example, graduate schools in information and library science are increasingly offering students the option to specialize in data curation. At the same time, domain scientists are struggling to find the right balance between data management and the research skills and disciplinary knowledge the next generation of researchers need to advance their respective fields. These issues raise the following questions:

- *How does the evolving nature of the data professions change education in the domain sciences and in information science, computer science and archives?*
- *What are the roles and responsibilities of the various professionals involved in data management?*

More Digital Data Products

The third way in which open access to data affects scale is through the potential to create new types of digital products that include data. These products might take a seemingly endless array of forms (e.g., artistic creations, educational learning modules, integrated data sets, end-to-end connections between laboratory results and published findings) that result from linking, integrating or weaving together data with publications, other data or other digital resources. The possibilities for these new products, as well as the challenges they present in terms of intellectual property, standardization and preservation, have been discussed by other authors (e.g., Borgman, [2007](#); Hey & Trefethen, [2008](#); NRC, [2010](#)). While creating new digital products that include data are not yet common, we expect this to change as technologies and policies evolve to support it; this will lead to new questions for researchers. Clearly, this is a wide-ranging and complex area, and the questions below only begin to outline the potential issues.

- 
- *What obstacles (e.g., legal, technical, social) exist to the creation of new digital objects and services that include data?*
 - *How do people create, share and reuse these new objects?*
 - *Who will be responsible for creating digital data objects that meet the needs of different types of reusers?*
 - *To what extent should these digital data products be captured and preserved?*
 - *What value do new products have beyond the components that were used to create them?*

Conclusions

The need to share and reuse data is an important topic in almost every high-level report or discussion concerning contemporary science. There are two overarching reasons for this emphasis. First, there is a belief that these activities are necessary to advance scientific research and solve important global problems. Second, there is a move to make the products of research available to a broad audience to support transparency, participation in the scientific process, and decision-making. Sharing and reuse at a large scale and over the long term imply that data can be accessed by anyone, anytime, from anywhere and used for any purpose, and that they can exist beyond the lives of the technology and the people who produced them.

In this paper, we argued that a dramatic increase in the amount of data, while important, is only one factor prompting the need for new research on data sharing and reuse. In order to achieve the vision of open access to data, we proposed a research agenda that addresses questions related to three other issues of scale: broader participation in data sharing and reuse, increases in the number and types of intermediaries, and more digital data products. Research in these areas should not only draw from and produce theory, but must also attempt to answer practical questions of curation for open access. Below we discuss three of the potentially many ways in which studies designed to address the research questions posed in this paper could inform data curation and the lifecycle model (Higgins, [2008](#)).

First, the challenges related to sharing data at a large scale and over the long term are new to many fields of science. Scientists in these areas are struggling with issues such as how to determine what data to save and for how long, how to ensure the accuracy and integrity of their own data as well as data they might reuse, how to document data for their own and others' use, and where to store data so they are broadly accessible. We believe an important first step toward supporting sharing and reuse can be gained through research that takes an in-depth look at how scientists manage data for their own use. This, coupled with more research on how people reuse data, would give data curators insight into how they might capitalize on scientists' personal documentation practices when creating documentation for others.

Second, it is often difficult to know in advance which data are valuable to curate and preserve. These decisions would become easier if we knew more about potential reusers of data: Who are they? What data skills do they possess? How do they find data and what do they need to know to reuse them? What are they interested in reusing (e.g., raw data, interpreted results, techniques)? Answers to these questions may help data curators better align their practices, policies and expertise in developing metadata standards, data file formats for preservation, and software, services and tools with the

needs of potential reusers. The answers may also inform the description and representation of information associated with data, and the capture, storage and dissemination of data.

Third, there are a growing number of choices available to people who want to deposit data. Yet, more options do not make it easier for people to select the best venue for sharing. Research that examines what influences how people make their data available may give data curators a better sense for what data are being placed where and the rationale behind these decisions. For instance, knowing the degree to which factors such as mandates from funding agencies, the existence of trusted repositories and supportive tools, and data producers' confidence in their data management skills may help data curators shape and advertise their capabilities, collections and services.


To conclude, our examples above illustrate how researchers can inform practice, but we also believe practitioners can and should inform the research. More active collaborations between the two are needed, such as joint formulation of research questions and proposals. Similarly, the research agenda we propose would benefit from collaborations among researchers from multiple disciplines, such as archives and computer, social, library and information, and domain sciences.

Acknowledgements

We appreciate comments from Nancy McGovern and three anonymous reviewers.


References

- Association of Research Libraries. (2006). *To stand the test of time: Long-term stewardship of digital data in science and engineering*. Washington, DC: Association of Research Libraries.
- Berg, M., & Goorman, E. (1999). The contextual nature of medical information. *International Journal of Medical Informatics* 56, (51-6). Shannon, Ireland: Elsevier Science.
- Birnholtz, J.P., & Bietz, M.J. (2003). Data at work: Supporting sharing in science and engineering. *Proceedings of the International ACM SIGGROUP Conference on Supporting Group Work, Sanibel Island, FL*. New York: ACM Press.
- Borgman, C.L. (2007). *Scholarship in the digital age: Information, infrastructure, and the Internet*. Cambridge, MA: MIT Press.
- Borgman, C.L., Wallis, J.C., & Enyedy, N. (2007). Little science confronts the data deluge: Habitat ecology, embedded sensor networks, and digital libraries. *International Journal on Digital Libraries* 7, (1-2) Berlin: Springer.
- Carlson, S., & Anderson, B. (2007). What are data? The many kinds of data and their implications for reuse. *Journal of Computer-Mediated Communication* 12, (2). Retrieved July 31, 2010, from <http://jcmc.indiana.edu/vol12/issue2/carlson.html>.

- 
- Collins, H.M. (1992). *Changing order: Replication and induction in scientific practice*. Chicago, IL: University of Chicago Press.
- Cragin, M.H., & Shankar, K. (2006). Scientific data collections and distributed collective practice. *Computer Supported Cooperative Work 15*, (2-3). Dordrecht, The Netherlands: Kluwer.
- Crow, R. (2002). *The case for institutional repositories: A SPARC position paper*. Washington, DC: The Scholarly Publishing & Academic Resources Coalition. Retrieved August 2, 2010, from http://scholarship.utm.edu/20/1/SPARC_102.pdf.
- Cummings, J., Finholt, T., Foster, I., Kesselman, C., & Lawrence, K. A. (2008). *Beyond being there: A blueprint for advancing the design, development, and evaluation of virtual organizations*. Arlington, VA: National Science Foundation.
- De Roure, D., Goble, C., Bhagat, J., et al. (2008). myExperiment: Defining the social virtual research environment. *Proceedings of the 4th IEEE International Conference on e-Science*. Indianapolis, Indiana.
- Faniel, I.M., & Jacobsen, T.E. (2010). Reusing scientific data: How earthquake engineering researchers assess the reusability of colleagues' data. *Computer Supported Cooperative Work 19*, (3-4). Dordrecht, The Netherlands: Kluwer.
- Green, A.G., & Gutmann, M.P. (2007). Building partnerships among social science researchers, institution-based repositories and domain specific data archives. *OCLC Systems & Services 20*, (1). Westport, CT: Meckler.
- Griffiths, A. (2008). The publication of research data: Researcher attitudes and behaviour. *International Journal of Digital Curation 4*, (1).
- Hesse, B. W., Sproull, L. S., Kiesler, S. B., & Walsh, J. P. (1993). Returns to science: Computer networks in oceanography. *Communications of the ACM 36*, (8). New York: ACM Press.
- Hey, T., Tansley, S., & Tolle, K. (2009). *The fourth paradigm: Data-intensive scientific discovery*. Microsoft Research. Retrieved August 2, 2010, from <http://research.microsoft.com/en-us/collaboration/fourthparadigm/default.aspx>.
- Hey, T., Trefethen, A. (2003). The data deluge: An e-science perspective. In F. Berman, G.C. Fox, & T. Hey, (Eds.), *Grid computing: Making the global infrastructure a reality*. New York: Wiley.



- Hey, T., & Trefethen, A. (2008). E-science, cyberinfrastructure, and scholarly communication. In G.M. Olson, A. Zimmerman, & N. Bos, (Eds.), *Scientific collaboration on the Internet*. Cambridge, MA: MIT Press.
- Higgins, S. (2008). The DCC Curation Lifecycle Model. *International Journal of Digital Curation* 3, (1).
- Jirotko, M., Procter, R., Hartswood, M., Slack, R., Simpson, A., Coopmans, C., Hinds, C., & Voss, A. (2005). Collaboration and trust in healthcare innovation: The eDiaMoND case study. *Computer Supported Cooperative Work* 14. Dordrecht, The Netherlands: Kluwer.
- Klein, J.T. (1996). Interdisciplinary needs: The current context. *Library Trends* 45, (2). Baltimore, MD: The Johns Hopkins University Press.
- Luther, K., Counts, S., Stecher, K. B., Hoff, A., & Johns, P. (2009). Pathfinder: An online collaboration environment for citizen scientists. *Proceedings of the 27th International Conference on Human Factors in Computing Systems*. New York: ACM Press.
- Markus, M.L. (2001). Toward a theory of knowledge reuse: Types of knowledge reuse situations and factors in reuse success. *Journal of Management Information Systems* 18, (1). Armonk, NY: M. E. Sharpe.
- National Research Council. (2010). *Steps toward large-scale data integration in the sciences: Summary of a workshop*. Washington, DC: National Academies Press. Retrieved August 2, 2010, from http://www.nap.edu/catalog.php?record_id=12916.
- National Science Board. (2005). *Long-lived data collections: Enabling research and education in the 21st Century*. National Science Foundation: Arlington, VA.
- Palmer, C.L. (1996). Information work at the boundaries of science: Linking library services to research practices. *Library Trends* 45, (2). Baltimore, MD: The Johns Hopkins University Press.
- Pierce, S.J. (1999). Boundary crossing in research literatures as a means of interdisciplinary information transfer. *Journal of the American Society for Information Science* 50, (3). New York: John Wiley & Sons.
- Shankar, K. (2007). Order from chaos: The poetics and pragmatics of scientific record keeping. *Journal of the American Society for Information Science and Technology* 58, (10). New York: John Wiley & Sons.

- 
- Van House, N.A. (2002). Digital libraries and practices of trust: Networked biodiversity information. *Social Epistemology 16, (1)*. London: Taylor & Francis.
- Van House, N.A. (2003). Digital libraries and collaborative knowledge construction. In A. P. Bishop, B. Bittenfield, & N. A. Van House, (Eds.), *Digital library use: Social practice in design and evaluation*. Cambridge, MA: MIT Press.
- Van House, N.A., Butler, M.H., & Schiff, L.R. (1998). Cooperative knowledge work and practices of trust: Sharing environmental planning data sets. *ACM Conference on Computer Supported Cooperative Work*. New York: ACM Press.
- Weeks, P., & Packard, J.M. (1997). Acceptance of scientific management by natural resource dependent communities. *Conservation Biology 11,(1)*. Boston, MA: Wiley-Blackwell.
- Whitfield, J. (2001). The budding amateurs. *Nature, 414, 578-579*. London: Macmillan.
- Zimmerman, A. (2007). Not by metadata alone: The use of diverse forms of knowledge to locate data for reuse. *International Journal on Digital Libraries 7, (1-2)*. Berlin: Springer.
- Zimmerman, A. (2008). New knowledge from old data: The role of standards in the sharing and reuse of ecological data. *Science, Technology, & Human Values 33, (5)*. London: Sage.