ASIST (American Society of Information Science and Technology)
Proceedings, Bringing Research and Practice Together
28Oct-02Nov 2005. Charlotte, North Carolina.

Interoperability Strategies for Scientific Cyberinfrastructure: Research and Practice

Karen S. Baker*
University of California San Diego, La Jolla, California 92093-0218
*corresponding author

David Ribes
University of California San Diego, La Jolla, California 92093-0533

Florence Millerand
University of California San Diego, La Jolla, California 92093-0092

Geoffrey C. Bowker
Santa Clara University, Santa Clara, California CA 95053-0470

The development of new infrastructures for research and collaboration are occurring together with changes in expectations for scientific knowledge. At differing rates, new vocabularies and perspectives are developing along with changes in the social and organizational practice of science. Between the new infrastructures and the new perspectives, we are changing both how we know and what it is to know.

A recently initiated three year project supported by the NSF Human Social Dynamics Program (Interoperability Strategies for Scientific Cyberinfrastructure: A Comparative Study) brings together work with three established research collaborations on large-scale information infrastructures in an attempt to understand the particular configurations of technologies, communities, and organizations. Despite specific configurations of technical commitment, community involvement and organizational structure, all the projects fall under a common rubric of achieving data interoperability.

Background

The historical notion of infrastructure as a background technical service has ramifications in research agendas today that call for interoperability of both community data efforts and scientific partnership (Star and Bowker, 2002). The goal of our ongoing study is to develop a coherent framework of analysis for large scale cyberinfrastructure development, with particular focus on interoperability mediation and achievement. Achieving data interoperability is understood here as a *process* that aims at information assemblage and sharing in order to enable reuse by various people, across diverging disciplines and across long periods of time.

A turn to interdisciplinary science in the 1950's was precipitated by the recognition that the complex theoretical and policy questions before scientists could not be answered within a single discipline using traditional scientific information practices. Traditional practices frequently have tied data to a particular investigator or project. Indeed, achieving data interoperability while often viewed as a technical issue – in terms of standards choices –unfolds into a more complex concern in the face of bringing in communities, and the practicalities of human organization or technical enactment.

While the framing of interoperability has remained primarily a technical one, it is with only a little prodding that our informants themselves come to see their work as thoroughly heterogeneous. Interoperability includes a broad range of work, including coordinating collaborations between diverse expert communities, building consensus on technical directions, securing consent from the domain, aligning interfaces with already existing community practices or training user populations. This kind of work is common to even the most technically centered interoperability endeavor, but it remains difficult to credit this crucial work and often remains beyond discursive capacity. Our research seeks to provide the conceptual, rhetorical and organizational tools to facilitate the work that is ongoing. For this approach, our research team necessarily consists of participants from Science and Technology Studies, Communication, Sociology, Information Management, and Social Informatics. Furthermore, this research depends on working in close relations with scientific community participants over time.

Methods and Approach

Our study draws from grounded theory building making use of ethnographic methods such as interviews, document analysis, participant observation, and community collaboration in order to develop a comparative perspective through cross case analysis (Strauss, 1987). By bringing the

rich ethnographic detail of the individual case studies together under a qualitative framework, we seek to gain an in-depth knowledge of the process of building scientific cyberinfrastructure.

The work of interoperability cannot be identified solely by pointing to a technology (e.g. metadata) or a goal (e.g. visualization). Rather, a consideration of a particular interoperability goal, say long-term data storage, *is also a consideration of* the durability of storage media, resources for data management, the stability of funding sources as well as of the willingness of scientists to invest their time in specifying the relevant data, adding metadata, and occasionally returning to update the archive. We seek to follow the entire range of heterogeneous action involved in what we call an *interoperability strategy*.

We roughly divide the strategies of interoperability into the a) data sharing and technology enactment; b) community features and community building approaches; and c) organization building within which technology is enacted and community mobilized. A descriptive comparative approach allows us to understand these interoperability strategies in their contexts as well as to discern similarities between the organizational rearrangements that may occur.

Case Studies

In this project, we are examining three interdisciplinary projects with quite different scientific goals and structures. Each is embedded in and aware of today's context of cyberinfrastructure and e-Science. GEON, the geo-sciences network, spearheaded by a national technology center, seeks to federate knowledge among all the various disciplines that make up the geosciences (Keller, 2003). The Long Term Ecological Research program is a federated network working to establish very long baselines of ecological data, beyond the scope of a single career (Hobbie et al., 2003). The incipient Ocean Informatics project is an informatics environment supporting earth system scientists and focusing on involvement of the scientific community in the design process (Baker et al, 2005).

This study will develop a grounded understanding of the organizational complexity in producing shared scientific cyberinfrastructure through assessment of three interoperability approaches developing within the three case studies: metadata standards, community ontologies, and local learning environments. In our research, we examine the distribution of analytic categories between the technical, the social and the organizational. When is a decision made to hardwire a protocol rather than to leave a verbal understanding in place? What kinds of organizational configurations are coupled with the creation of a federated database? Through comparison, we aim to consider a unified framework for describing fundamental dimensions of interoperability and cyberinfratructuring.

Building from the salient features of infrastructure (Star and Ruhleder, 1996) and collaboration (Olson and Olson, 2000; Finholt 2000), we will examine community characteristics in general and coordination mechanisms in particular. Communities incorporate, just as information systems instantiate, both tacit and explicit methods, values, epistemologies and ontologies. Our cross-community work helps identify diverse capabilities and choices. We expect this will illuminate competing values, models, and mindsets in addition to fostering awareness of different discourses, value orientations, and cultural prejudices. We seek an expanded vocabulary for interoperability of data and communities as well as for the roles of technology and participants supporting this interoperability.

Our main hypothesis is that interoperability endeavors come along with - as they both require and provoke - organizational and inter-organizational redefinitions and rearrangements, which bring several challenges (e.g. new attitudes toward information, re-distribution of power structures, new organizational forms, re-definition of territories…). Before being materialized in technical artifacts and organizational structures, interoperability strategies develop in the interplay of *negotiations*. Using ethnographic methods, our empirical work will focus precisely on the study of those negotiations.

Conclusion

The work of achieving interoperability is always-already a work of reorganization, community building, and decision making. Our research does not seek to produce a new sphere of 'social action' within cyberinfrastructure, but rather to identify and enable action within already existing socio-technical work. In identifying a fuller range of effect in building cyberinfrastructure, we are hoping to contribute to a considered design of these broader ramifications. A correlate of this model is to produce a larger vision of cyberinfrastructure, a recognition that a domain cannot merely outsource its technical infrastructure without also outsourcing its organization, community relations and possibly even its epistemology (Fountain, 2001).

The aim is an integrated understanding of interoperability as a complex planning-and-action, a linking of 'social' and 'technical' strategies. There is currently a large gap by implementers between design and assessment with enactment of interoperability because the majority of explicit attention is focused on a technical implementation – getting the system designed and operational. Rather than advocating to redistribute this economy of attention, the *specific* goal of this project is a deep tying-together of *particular* technologies of interoperability with a *particular* set of human/organizational concerns. The larger goal of

the project is providing an overarching framework in which it is *understood* that in considering interoperability, one is considering a tied sociotechnical approach. This tying together may be referred to as an 'interoperability strategy' where the links between technical choice and social organization are explicitly recognized.

ACKNOWLEDGEMENTS

This work is supported by an NSF/SBE/SES Human Social Dynamics grant #04-33369 (Interoperability Strategies for Scientific Cyberinfrastructure: A Comparative Study) as well as Social Sciences and Humanities Research Council of Canada Postdoctoral Fellowship #756-2003-0099. The work is conducted in collaboration with GEON and LTER communities (NSF/OCE #04-17616, NSF/OPP #02-17282 and #04-05069) as well as the Ocean Informatics Environment supported at Scripps Institution of Oceanography and the Laboratory of Comparative Human Cognition.

REFERENCES

Baker K.S., Jackson S.J., & Wanetick J.R. (2005). Strategies Supporting Heterogeneous Data and Interdisciplinary Collaboration: Towards an Ocean Informatics Environment. Proceedings of the Hawaii International Conference for System Science, HICSS38.

Finholt, T.A. (2004). Collaboratories. EB Cronin (ed) Annual Review of Information Science and Technology.

Fountain, J. (2001). Building the Virtual State: Information Technology and Institutional Change. Brooking Institution Press, Washington, DC.

Keller, R.B. (2003) GEON (GEOscience Network)-A First Step In Creating Cyberinfrastructure for The Geosciences. Electronic Seismologist, July/August 2003.

Hobbie, J.E., Carpenter, S.R., Grimm, N.B., Gosz, J.R., & Seastedt, T.R. (2003). The US Long Term Ecological Research Program. BioScience 53(2), 21-32.

Olson, G.M. & Olson, J.S. (2000). Distance Matters. Human-Computer Interaction 15, 139-178.

Star, S.L. & Ruhleder, K. (1996). Steps Toward an Ecology of Infrastructure: Design and Access for Large Information Spaces. Information Systems Research 7(1), 111-134.

Star, S.L. & Bowker, G. (2002). How to Infrastructure *In* Lievrouw and Livingstone (Eds.), The Handbook of New Media (pp. 151-162). London: SAGE Publications.

Strauss A (1987). Qualitative Analysis for Social Scientists. Cambridge University Press, Cambridge.