Comparative Interoperability Project: Configurations of Community, Technology, Organization

David Ribes[1], Karen S. Baker[2], Florence Millerand[3], and Geoffrey C. Bowker[4]

[1] Sociology/Science Studies
University of California,
San Diego
La Jolla, CA 92093, USA
1.858.534.4627
dribes@ucsd.edu

[2]Scripps Institution
of Oceanography
University of California,
San Diego
La Jolla, CA 92093-0218,
USA
1.858.534.2350
kbaker@ucsd.edu

[3] LCHC/Science Studies
University of California,
San Diego
La Jolla, CA 92093, USA
1.858.534.4006
fmillera@ucsd.edu

[4]Ctr.f/ Science Technology
& Society
Santa Clara University
Santa Clara, CA USA
1.408.551.6058
gbowker@scu.edu

ABSTRACT

In this paper we describe the methods, goals and early findings of the research endeavor 'Comparative Interoperability Project' (CIP). The CIP is an extended interdisciplinary collaboration of information and social scientists with the shared goal of understanding the diverse range of interoperability strategies within information infrastructure building activities. We take interoperability strategies to be the simultaneous mobilization of community, organizational and technical resources to enable data integration. The CIP draws together work with three ongoing collaborative scientific projects (GEON, LTER, Ocean Informatics) that are building information infrastructures for the natural sciences.

Categories and Subject Descriptors

K.4.3 [Organizational Impacts]: *Computer-supported collaborative work*

General Terms

Design, Human Factors, Standardization, Theory

Keywords

interoperability, organization, community, infrastructure,CSCW

1. INTRODUCTION

In this paper we describe the efforts of the 'Comparative Interoperability Project' (CIP) to bring together insights of three ongoing projects within the social study of information infrastructures. The goal of the project is to understand the simultaneous mobilization of community, organizational and technical resources for enabling data integration. Within the natural sciences there has been a progressively stronger call for the integration of data across traditional disciplinary boundaries, but the strategies of data interoperability remain largely unarticulated and poorly understood. The technologies themselves are novel, and new forms continue to emerge. Furthermore, while the technologies of interoperability are commonly the initial focus, it often emerges that community mobilization is a significant challenge within interoperability efforts. The CIP brings together and explores the insights of three ongoing projects in order to foster an understanding of links between community, technology and data interoperability. In this paper we present the goals, methods and initial framing for an interdisciplinary team working together with science communities.

Data interoperability is a form of infrastructure [1]. By this we mean that it is not a tool for a single scientist or even a research team, rather it is an investment intended to serve as a long-term resource for a broader community. Within information infrastructure projects interoperability is often defined as the common goal of a collective [2]. The three studied communities within CIP are GEON, Long-Term Ecological Research, and Ocean Informatics. All three projects have social science researchers as participants and/or observers. We have chosen to bring together these particular projects for their shared interest in data-integration, and to learn from their diversity of approaches in achieving that integration.

GEON, the geo-sciences network (geongrid.org), is a five year cyberinfrastructure project with the goal of providing computing resources, data integration and mapping/visualization tools for the broader solid-earth sciences. GEON itself is a nationally distributed project, with nodes spread across the US, while its technical core is centered at the San Diego Supercomputer Center (SDSC). GEON is funded 'from above' by the NSF, but is driven 'from below' by geo-scientists and their information technology collaborators at the SDSC.

Long-Term Ecological Research (LTER; lternet.edu) is a thirty year program that brings together diverse ecological sciences for the purpose of enabling interdisciplinary collaboration and producing data integration matching ecological timespans [7]. In recent years the social, organizational and technical complexities of ensuring data interoperability across time and specialties has led to a greater formalization of data integration efforts.

Finally, Ocean Informatics (OI) is a nascent information infrastructure for the ocean sciences centered at the Scripps Institution of Oceanography [3]. OI has taken an inductive approach to the question of interoperability by first exploring and articulating an informatics conceptual framework; by taking careful survey of the local community needs and available resources; and by investing in community enrollment before beginning large-scale data integration efforts.

The three projects share the common goals of data integration across traditional disciplinary lines, but have chosen diverse *interoperability strategies* of community mobilization, technical direction and organizational structure.

JCDL'05, June 7–11, 2005, Denver, Colorado, USA

2. RESEARCH AND FINDINGS

The CIP employs methodologies derived from sociology, social informatics [4] and science and technology studiess (STS). The research focus is on the production of detailed qualitative case studies and cross-case analysis. Data collection methods consist of primary field research – which includes ethnographic participant observation, coupled with selective interviews – and archiving of secondary materials such technical documentation and research output. The activities within GEON, LTER and OI are highly heterogeneous, and the methods of this study are tailored to make possible the following of these diverse actions [5].

In a single day an information 'technologist' may shift from designing a detailed technical protocol to consulting with earth scientists about the protocol, and then to writing a report about the protocol for a newsletter or technical bulletin. It is this heterogeneity of activities by scientists and information technologists themselves to which we call attention. By following a practitioner across task boundaries usually kept separate – such as 'science', 'communicating with the community' and 'writing code' – we form a broader image of the work involved in achieving interoperability. In each of the three examples below we draw attention to a relation between the particular characteristics of a domain community, the technological trajectory and the organizational action which links them:

i) one particular strategy within GEON has been the creation of ontologies for data integration and knowledge mediation. Ontologies are formal conceptual maps of domain knowledge. By tying ontologies to datasets, or subsections of data, the user is able to navigate with greater ease across unfamiliar databases or knowledge domains. In building ontologies there is the two part difficulty of i) specifying domain knowledge and ii) then communicating this for information technologists to represent in machine language. The organizational solution within GEON has been 'ontology workshops' – small groups of geo-scientists and ontology experts brought together to work-out particular ontologies. This foregrounded technical work is coupled with community outreach efforts to ensure consent and awareness of the new ontology resources for the broader geo-science community [6].

ii) LTER has endorsed a community metadata standard: the ecological metadata language (EML). Metadata is 'data about data'; fine-grained detail about the structure and content of data can facilitate interoperability. But the existence of a metadata standard is only useful if implemented and maintained. For broad-scale success individual ecological scientists must describe their data in EML. This involves a significant investment of time which does not immediately benefit the researcher. Even with a well intentioned or 'incentivized' researcher, there remains the work of technical mediation: a deep familiarity and upkeep with EML itself. The success of metadata standards relies on a mobilization of communities of scientific practitioners which requires continuous individual investment. Thus the success of EML – achieving interoperability – is also a transformation of the daily practices and organization of ecological scientists.

iii) within Ocean Informatics, information managers have worked closely with ocean scientists to make informed decisions about what kinds of organizational work are coupled to technical choice. By drawing together experience from previous technical efforts and research from social science studies of environment-science communities, a body of knowledge about sociotechnical change is made available thus opening the possibility for reflexive community participation in their own transformation prior to technical implementation.

4. CONCLUSION

In this paper we have introduced the Comparative Interoperability Project which brings together insights of three ongoing projects within social informatics, and seeks to understand the tied configurations of technical approach, community mobilization, and organizational structure.

While interoperability has been treated primarily as a technical concept, our initial research has shown that in practice interoperability involves the alignment of technologies, community mobilization and organizational structure. It is with only a little prodding on our part that technical practitioners themselves come to see their work as thoroughly heterogeneous: from coordinating and facilitating collaborations between diverse expert communities; building consensus on technical decisions and the future investment of work; aligning interfaces with already existing community practices or; training user populations. The CIP seeks to contribute to development of a vocabulary of action, and organizational resources for coordinating that action, which matches the heterogeneity and sophistication of the practices involved in producing data interoperability.

5. ACKNOWLEDGEMENTS

This work is supported by an NSF/SBE/SES Human Social Dynamics grant #04-33369 (Interoperability Strategies for Scientific Cyberinfrastructure: A Comparative Study), SSHRC Postdoctoral Fellowship #756-2003-0099, and works in synergy with GEON and LTER communities (GEON ITR (EAR)-0225673, NSF/OCE #04-17616, NSF/OPP #02-17282 and OPP/SBE/#04-05069) as well as the Ocean Informatics Environment supported at Scripps Institution of Oceanography.

6. REFERENCES

[1] Star, S.L. and G.C. Bowker, *How To Infrastructure*, in *The Handbook of New Media*, L.A. Lievrouw and S.L. Livingston, Editors. 2002, SAGE Publicaitons: London. p. 151-162.

[2] Kling, R. and W. Scacchi, *The Web of Computing: Computer Technology as Social Organization.* Advances in Computers, 1982. **21**: p. 1-90.

[3] Baker, K.S., S.J. Jackson, and J.R. Wanetick, *Strategies Supporting Heterogeneous Data and Interdisciplinary Collabortion: Towards an Ocean Informatics Environment.* Proceedings of the Hawaii International Conference for System Science, HICSS38, 2005.

[4] Kling, R., *What is Social Informatics and Why Does it Matter?* D-Lib Magazine, 1999. **5**(1): p. 1-23.

[5] Latour, B., *Science in action: how to follow scientists and engineers through society*. 1987, Cambridge, Mass.: Harvard University Press. 274.

[6] Ribes, D. and G.C. Bowker, *Ontologies and the Machinery of Difference.* Journal for the Association of Information Systems, forthcoming.

[7] Hobbie, J.E., S. R. Carpenter, N. B. Grimm, J. R. Gosz, & T. R. Seastedt, *The US Long Term Ecological Research Program*. BioScience 53(1), pp. 21-32 (2003).