

Scientific Infrastructure Design: Information Environments and Knowledge Provinces

Karen S. Baker (corresponding author)

Scripps Institution of Oceanography, University of California at San Diego (UCSD), La Jolla, CA 92093-0218;
Tel: 858-534-2358 kbaker@ucsd.edu

Florence Millerand

Department of social and public communication, Université du Québec à Montréal (UQAM), Montreal, QC
H3C 3P8, CANADA; Tel: 514-987-3000 3593# millerand.florence@uqam.ca

Conceptual models and design processes shape the practice of information infrastructure building in the sciences. We consider two distinct perspectives: (i) a cyber view of disintermediation where information technology enables data flow from the ‘field’ and on to the digital doorstep of the general end-user, and (ii) an intermediated view with bidirectional communications where local participants act as mediators within an information environment. Drawing from the literatures of information systems and science studies, we argue that differences in conceptual models have critical implications for users and their working environments. While the cyber view is receiving a lot of attention in current scientific efforts, highlighting the multiplicity of knowledge provinces with their respective worldviews opens up understandings of sociotechnical design processes and of knowledge work. The concept of a range of knowledge provinces enables description of dynamic configurations with shifting boundaries and supports planning for a diversity of arrangements across the digital landscape.

Introduction

As digital networks and ‘web 2.0’ become well-recognized terms in our age of digital information, they influence design. They represent conceptual models embedded within our understandings of digital services and of data users. As their use becomes more pervasive, the terms become more known yet less visible. Drawing from the literatures of information systems and science studies, particularly works in the emergent field of infrastructure studies, we focus on the diversity of arrangements associated with data use and data users as critical to the design of scientific information infrastructure.

Cyberinfrastructure initiatives have developed in recent years, particularly in the natural sciences, to involve large-scale communities and information systems. Increasing technological capacities enhance data-taking capabilities and are inspiring plans to instrument the landscape. In considering how we think about information infrastructure for the sciences, the organization of knowledge provinces appears as an element in the digital landscape, one that shapes scientific services to multiple users and influences data flow.

Key to infrastructure understandings is the notion of mediation (e.g. Latour, 1991, 1994; Friedman, 1989); intermediation and disintermediation are two design processes with differing views of users. We use intermediation to refer to the relation between either an individual and other participants or objects such as datasets or information systems. Disintermediation in the business arena refers to eliminating traditional intermediaries. Within information science, it refers to the act of bypassing information intermediaries such as librarians with automated information retrieval systems or data managers with automated delivery systems. The implications of disintermediation for practitioners vary case by case, e.g. for library users when library practices shift or for scientists when data practices change. *We consider mediation approaches and their consequences in the practice of information infrastructure design and development.*

Two Views

A common conceptual model for data transfer is a pipeline (Figure 1a) where focus is on data flow from 'the field' through a center to a general end-user. The view is one of disintermediation with information technology enabling data streams delivered routinely and anonymously to general end-users. End-users include scientists (the ones who collect the data as well as the ones who make use of it e.g. a scientist from another discipline), technologists, educators, and even people at large. This technologically ideal scenario carries with it an assumption that more data and more data access result in more knowledge. Data centers, frequently staffed by individuals trained in technology and information management, optimize data processes for data preservation and data access.

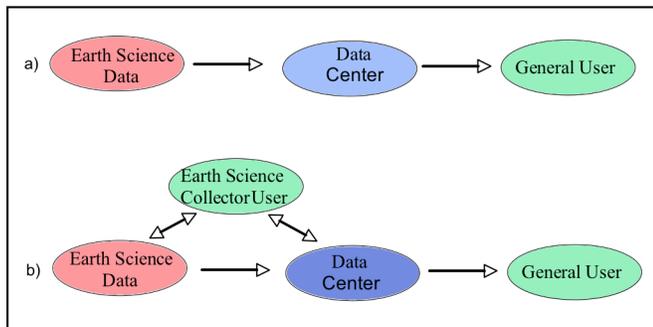


Figure 1. Information environment scenarios: a) disintermediation for a global general user; b) intermediation by the earth science data designer-collector-user teams constituting a local knowledge province that supports local science users as well as data delivery to a general end-user.

Cyberinfrastructure (CI), a buzzword today, is frequently associated with pipelines and grids. Defined in the technical realm as residing between the layer of base technology and specific use, there is “the cyberinfrastructure layer of enabling hardware, algorithms, software, communications, institution, and personnel” (Atkins, 2003).

We distinguish three understandings of CI: a) a solution, b) a growth option and/or c) one of a plurality of modular elements that together represent a federated scenario. In the first case, cyberinfrastructure may be seen as providing a general mechanism enabling global information flows, that is, an upgrade to today’s independent data centers. In this view, the aim is to align a single functional cyberinfrastructure. Another view is to consider cyberinfrastructure as growing over time, where initial, simple approaches are left behind as we transition to a new CI solution. Claims of progress frequently attend both these cases. The promise of automated data streams coupled with replacement or augmented solutions, leads to streamlined models and “cyber assumptions” with respect to pipelines and disintermediation.

While the first two understandings (a solution or a growth option) are prevalent, we argue that understanding CI as an element of a federated scenario allows for grasping the dynamics of CI as complex sociotechnical systems (Figure 1b). With this approach the earth scientist is brought into the picture; the earth scientist role involves a team - data collector, data scientist, analyst, manager, and technologist – working together on measurement design as well as data collection, processing, and use. As members of a field-connected workplace sub system, they are local data mediators connecting the organizational and technological to field practices and experience. In also linking local to larger-scale or down-stream needs, the work constitutes a rich knowledge province focused on the data, it’s quality and local use. In the local context, the data use and management for scientific research provide services such as quality control, analysis, and attention to changes that contribute to sustainability of networked data flows.

Analogous to the layering metaphor used in computer science that describes middleware as software that connects separate applications between layers and across networks, a local data team represents a ‘middlewhere’, instantiating a knowledge-creating province that provides evolving services in relation to the individual dataset on one hand and the networked collection on the other.

Who Are the Users?

A pipeline or disintermediation scenario is framed as an automated technical procedure designed for optimized data flow from the field. Associated with cyberinfrastructure are conceptual models of users in terms of a general end-user and infrastructure design processes in terms of universal systems for large-scale collaboration. Delivery is to a global end-user. Though content is local, it is wrapped with descriptive metadata providing contextualization. Local participants with close ties to the data deliver electronic bits but also participate as both data takers and data users who possess an indigenous knowledge of the data and its context (Millerand and Baker, accepted).

A local information environment is a generative forum for dialogue about the data where participants create, support, and sustain communicative functions related to data and its flow. Communication and information systems work includes data handling such as quality control and structuring, information elicitation such as requirements analysis and metadata creation as well as activities such as design, articulation, and enactment (Suchman, 1991; Hirschheim and Klein, 2003; Bratteteig, 2003; Baker and Millerand, 2007). Local users, as mediators, perform as bidirectional interfaces addressing both local issues as well as facilitators who carry out preparations needed for 'passing the data forward' in federating networks. What is at stake is the amount, quality, and access to data. Local histories with regional specificity require underappreciated efforts by these participants to capture all that the information that would describe this data well. How to capture 'all the metadata' unfolds into a set of as yet ill-defined questions that are central to integration of data within local environments as well as across sites with attendant institutional boundaries. This work is the subject of active research in the fields of information science, information systems, and computer science on the one hand and informatics, information management and the particular domain science efforts on the other hand. This work is influenced and informed by a host of users. In adjusting fieldwork and data handling to be coordinated with information elicitation and systems design, local teams negotiate the balancing of intertwined social, organizational and technological arrangements that optimize data packaging for both local and general use. Building upon the notion of thick descriptions (Geertz, 1973) and thick things (Alder, 2007), an information environment may be described as "thick infrastructure" (Jackson and Baker, 2004) involving complex socialtechnical systems and social worlds (Kling and Lamb, 1999; Clarke 1991, 2006). Ethnographic studies of users and user communities with a variety of modes of engagement are providing insight into the multiple dimensions of scientific work and its relation to information infrastructure building (e.g. Star and Ruhleder 1996; Karasti and Baker, 2004; Baker et al, 2005; Lee et al, 2006; Ribes, 2006; Millerand and Bowker, forthcoming).

Interdependent Knowledge Provinces

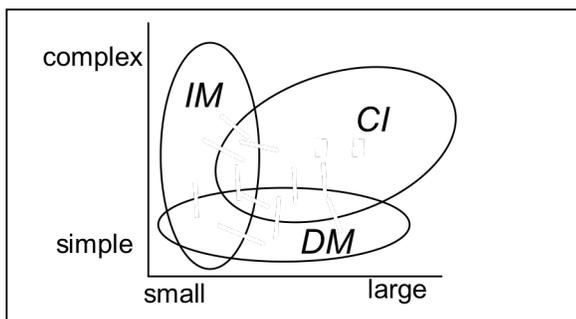


Figure 2. Knowledge provinces: a pluralistic view distinguishing interdependent knowledge provinces representing a multiplicity of work arenas including data management (DM), information management (IM) and cyberinfrastructure (CI).

Information infrastructures have been described as ways of knowing in a networked environment (Bowker et al, forthcoming). Designing network environments entails a number of interrelated work environment factors (or constraints) that shape work activities (Fidel, 2006). Scientific activities occur within heterogeneous and distributed teams, constantly transiting from local (the field and the lab) to global contexts (large-scale interdisciplinary projects). Thus, scientific work is supported by and benefits from a variety of types of data handling arenas.

Figure 2 provides a view framed in two dimensions: a small-large axis and a simple-complex axis. Distinct but complementary knowledge provinces are shown schematically within a landscape supported by information

infrastructures. The sizes and locations of the provinces will change depending upon circumstances. The data management province focuses on targeted data types (simple) and automated homogeneous data streams spanning small to large; the information management arena focuses on handling of multiple heterogeneous data types (small) and development of relations between them (simple to complex); and the cyber province scales into archive and access of large amounts of data as well as complex arrays of data streams typically associated with national centers. This conceptual model shows local data and information management provinces overlapping the cyberinfrastructure arena. Overlaps indicate critical integrative boundary interfaces; dashed lines suggest openness for information exchange. Such a federated landscape provides a mechanism for distinguishing provinces and prevents contemporary notions of large-scale cyberinfrastructure from marginalizing other provinces. Figure 2 provides a conceptual platform for pluralism. This coordinated set of work within provinces *and between them* has ramifications for data taking and local data use, information access and global data use.

Paying attention to the growth of local knowledge-making provinces is a strategy for changing how we think about generalizations and network federation. Considering local and global arenas as distinct yet interdependent, allows a local data center to be transformed, to be seen as a new-age information environment where design and communication efforts are understood as knowledge work. Such work includes concern with relations between the local and global as well as between multiple forms of knowledge (Hirschheim and Klein, 2003). The Atkins Report (2003) identifies the challenge as “one of design of knowledge environments for multiple uses” and states:

...the opportunity is here to create cyberinfrastructure that enables more ubiquitous, comprehensive knowledge environments that become functionally complete for specific research communities in terms of people, data, information, tools, and instruments, and that include unprecedented capacity for computational, storage, and communication”. (Atkins et al, 2003)

A subsequent report, the History and Theory of Infrastructure (Edwards et al, 2007), elaborates on the notion of infrastructure and addresses the dynamics of growing sociotechnical information infrastructure over time. This report cautions: “As multiple systems assemble into networks, and networks into webs or ‘internetworks,’ early choices constrain the options available for moving forward, creating what historical economists call ‘path dependence.’” And the report summarizes:

Speaking of cyberinfrastructure as a machine to be built or a technical system to be designed tends to downplay the importance of social, institutional, organizational, legal, cultural, and other non-technical problems developers always face...Hence this report turns away from a language of design and engineering, reframing the discussion in a more organic lexicon. Since infrastructures are incremental and modular, they are always constructed in many places (the local), combined and recombined (the modular), and they take on new meaning in both different times and spaces (the contextual). (Edwards et al, 2007)

Recognizing the need for relations between provinces with differing worldviews is an integrative effort that acknowledges local differences as well as insights from both local practices and national or international collections. It allows breaking out of local-global dichotomies and out of technology-science or economy-of-scale versus complexity-of-scale perspectives so that fuller understandings of knowledge types and categories, uses and users can emerge. Relations may be between multiple users (local and end-user, intermediate and end-user) or multiple knowledge provinces (data management, information management, and cyberinfrastructure). A key information age challenge is creating a modular system of decentralized, heterogeneous information environments that function as learning arenas across the digital landscape. Wherever a diversity of knowledge provinces is both recognized and sustained, negotiations are critical and ongoing between those working in different provinces who seek to move data and information across boundaries. Such work constitutes the art and science of contemporary information infrastructure building.

Acknowledgements

This work is supported by NSF HSD Grant #04-33369 and the Social Sciences and Humanities Research Council of Canada Postdoctoral Fellowship #756-2003-0099. The work is in collaboration with LTER (NSF/OCE #04-17616 and NSF/OPP #02-17282) and Ocean Informatics located at Scripps Institution of Oceanography.

References

- Alder, K., 2007. Thick Things: Introduction. *The History of Science Society* 98:80-83.
- Atkins, D. et al., 2003. NSF-AP Report: Revolutionizing science and engineering through cyberinfrastructure: Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure.
- Baker, K.S., Ribes, D., Millerand, F. & Bowker, G.C. (2005). Interoperability strategies scientific infrastructure: research and practice. *American Society for Information Systems and Technology Proceedings, ASIST 2005*.
- Baker, K.S. & Millerand, F. (2007). Articulation work supporting information infrastructure design: coordination, categorization, and assessment in practice. *Proceedings of the 40th Hawaii International Conference on System Sciences. HICSS40, IEEE Computer Society, New Brunswick, NJ*.
- Bowker, G.C., Baker, K.S., Millerand, F. & Ribes, D. (forthcoming). Towards information infrastructure studies: ways of knowing in a networked environment. In J. Hunsinger, M. Allen, L. Klasrup (Eds). *International Handbook of Internet Research*,
- Bratteteig, T. (2003). Making change: dealing with relations between design and use. PhD Thesis, University of Oslo.
- Clarke, A. E. (1991). Social worlds/arenas theory as organizational theory. In D. Maines (Ed), *Social organization and social processes: Essays in Honour of Anselm L. Strauss*, NY: Aldine Gruyter.
- Clarke, A.E. (2006). Social worlds. In George Ritzer (Ed.) *The Blackwell Encyclopedia of Sociology*, Vol. IX. Malden MA: Blackwell, 4547-9.
- Edwards, P.N., Jackson, S.J., Bowker, G.C., & Knobel, C. (2007). Understanding infrastructure: dynamics, tensions, and design, NSF Report of a Workshop: History and theory of infrastructure: lessons for new scientific cyberinfrastructures. <http://hdl.handle.net/2027.42/49353>.
- Fidel, R. (2006). An ecological approach to the design of information systems. *Bulletin of the ASIS&T* Oct/Nov, 6-8.
- Friedman, A. L. (1989). *Computer systems development: history, organization and implementation*. Chichester, UK: Wiley.
- Geertz, C. (1973). Thick description: toward an interpretative theory of culture. In *The Interpretation of cultures*. New York: Basic Books.
- Hirschheim, R. & Klein, H.K. (2003). Crisis in the information system field? A critical reflection on the state of the discipline. *Journal of the Association for Information Systems*. 4(5), 237-293.
- Jackson, S.J., & Baker, K.S. (2004). Ecological design, collaborative care, and ocean informatics. *Proceedings of the Participatory Design Conference, PDC-04, Vol 2*. Toronto, Canada.
- Karasti, H. & Baker, K. (2004). Infrastructuring for the long-term: ecological information management. *Proceedings of the 37th Hawaii International Conference on System Sciences. HICSS37, IEEE Computer Society*.
- Kling, R. & Lamb, R. (1999). IT and organizational change in digital economies: a socio-technical approach. *Computers and Society*, 29(3), 17-25.
- Millerand, F., & Baker, K.S. (forthcoming). Who are the users? From a single user to a web of users in the design of a working standard. *Information Systems Journal*.
- Latour, B. (1994). On technical mediation. *Philosophy, Sociology, Genealogy. Common Knowledge* 3 (1994), 29-64.
- Latour, B. (1991). *We have never been modern*. Cambridge, Harvard University Press.
- Lee, C.P., Dourish, P. & Mark, G. (2006). The Human infrastructure of cyberinfrastructure. *Computer Supported Cooperative Work Conference (CSCW)*, Banff, Canada.

- Millerand, F. and Bowker, G.C. (forthcoming). Metadata standards. trajectories and enactment in the life of an ontology. In S. L. Star & M. Lampland (Eds.), *Formalizing practices: reckoning with standards, numbers and models in science and everyday life*.
- Ribes, D. (2006). *Universal informatics: building cyberinfrastructure, interoperating the geosciences*. Department of Sociology (Science Studies). San Diego, University of California. Unpublished Ph.D. Dissertation.
- Star, S.L. & Ruhleder, K. (1996). Steps Towards an ecology of infrastructure: design and access for large information spaces. *Information Systems Research*, 7(1), 111-134.
- Suchman, L. (1991). Supporting articulation work. In *Computerization and controversy: value conflicts and social choices*, 407-423.