

## 17 Organizing for Multidisciplinary Collaboration: The Case of the Geosciences Network

David Ribes and Geoffrey C. Bowker

Within the sciences, infrastructure has come to mean much more than “tubes and wires.” Contemporary infrastructure-building projects for the sciences—often dubbed cyberinfrastructure—seek to develop the communication capacity to collaborate across distances and institutional barriers (Star and Ruhleder, 1994), work across the technical differences endemic to specialized disciplinary work, and manage the increasingly large and heterogeneous archives of scientific data (Bowker 2000). The goal of infrastructure building today is to encourage multiple configurations of collaboration and enable novel interdisciplinary research ties. Fostering such ties is no easy task. Developing collaborative ventures stretches well beyond the confines of the “technical” to addressing problems centrally defined as “sociological,” such as forming communities, communicating across disciplinary boundaries, or meeting the needs of diverse career reward systems.

In this chapter, we focus on the work of multidisciplinary participants as they went about planning and building the Geosciences Network (GEON). GEON, a cyberinfrastructure project, seeks to produce a repertoire of high-end information technologies for the broader earth sciences:

The ultimate goal of GEON is to establish a new informatics-based paradigm in the geosciences, to provide a holistic understanding of the Earth’s dynamic systems, thereby transforming the science. (GEON, 2002, 3)

GEON is intended to be an “umbrella infrastructure” for the geosciences, bringing together tools for collaboration and data that will serve the heterogeneous disciplines that study the earth. In organizing to produce this umbrella infrastructure, the participants drew together a wide range of earth and computer science experts representing multiple institutions across the United States. The network has twelve principal investigators (PIs) roughly split between those studying the earth and those studying computation, thereby presenting difficulties for collaboration. This is the first and most obvious disciplinary boundary to be crossed: computer and earth sciences. Yet there is also a second, often less prominently discussed, set of disciplinary boundaries: the

earth science members of GEON are themselves subdivided by their expertise, which includes paleobotany, geophysics, and other specialties; the disciplines of the earth sciences vary by method and focus.

Both axes of collaboration must be rendered explicit in order to understand the work of developing infrastructure within the model of cyberinfrastructure: first, collaboration is across the domains (domain/domain), and second, collaboration is between computer and domain scientists (computer science/domain).<sup>1</sup> Both axes require work to overcome communication and organizational barriers. In this chapter, we trace three temporal phases as GEON participants sought to cross *both disciplinary axes of difference*; we capture and articulate the tactics and strategies as they went about building an umbrella infrastructure bringing together the heterogeneous earth and computer sciences.

GEON is a project in motion. At the time of this writing, GEON remains at the prototype stage. Thus, we do not focus on the end-product infrastructure; instead we analyze the practical processes in building the infrastructure (Bowker and Star 1999)—this is to date typical for such studies, since few specifically cyberinfrastructure projects have built up an extensive user base. The plan for the chapter is as follows. We first outline the two axes of collaboration, noting the particular difficulties that arise at each. We then explore the tactics and strategies adopted by the GEON participants to address working within the geosciences, and collaborating across earth and computer sciences. We focus on three phases in the project's early development. Our empirical analysis begins "before" GEON—that is, during its proposal-writing stage. It is here that the participants negotiated a vision for multidisciplinary collaboration; in the case of GEON, we found that key notions for the project were articulated such as a "balance" between computer and earth sciences research. Second, we focus on the initial meetings of GEON: the "kickoff" and "all-hands." While in the proposal participants put forward a multidisciplinary vision for the earth sciences, actually building that collaboration was a practical activity. The initial meetings set aside considerable time to begin forming a "GEON community," as earth and computer scientists alike learned of their disciplinarily grounded differences. Finally, we focus on the empirical "test beds" that served to coordinate work across disciplinary boundaries. The test beds are of scientific interest to the multiple constituencies participating in this cyberinfrastructure venture: Over time, geologists in multiple domains have developed extensive knowledge about these areas; and from the perspective of computer science it is the data sets themselves that are of interest.

#### Notes on Method

From its formal inception in 2002 we were invited to participate in the GEON project as "social informatics researchers." The PIs of the team had themselves identified fu-

ture complications for working across computer science/domain and domain/domain boundaries: they recognized that communication across forms of expertise could be a bottleneck in the collaborative venture. The focus of this chapter reflects this interest on the part of the GEON participants. We characterize our research stance in this project as "social dimensions feedback" (Ribes and Baker 2007), in which our primary role as observers was coupled with occasional requests to communicate feedback and research findings. Our investigations have resulted in various opportunities to consult with GEON participants, the host institution (the San Diego Supercomputer Center) and the broader geoscience community.

The research was driven by grounded theory methodology (Clarke 2005; Star 1999): iterations of data collection were combined with *testing* against substantively generated theory as well as *constant comparisons* with historical and contemporary studies of infrastructure (Ribes and Finholt 2007). Between 2002 and 2005 we conducted ethnographic data collection, attending the meetings, workshops, and conferences organized within GEON. Such events were audio recorded, annotated, and selectively transcribed; archives were maintained using the qualitative analysis software suite NVivo. Furthermore, we were granted unconditional access to the various GEON e-mail Listservs, providing a voluminous and finely detailed stream of data. In the later years of data collection, the ethnographic research was supplemented by interviewing GEON participants, key members of the National Science Foundation (NSF), and representatives of the earth science institutions (such as the U.S. Geological Survey).

### Two Axes of Collaboration

We should not treat "multidisciplinary collaboration" as a homogeneous entity. The configuration of each collaboration is specific. Influences on the character of collaboration include the particular representation of domain participants, the length of the engagement, or the purposes for working together. We focus on two critical axes of collaboration in infrastructure-building projects: multidisciplinary relations in GEON are *across the geosciences (domain/domain)* and *with computer scientists (computer science/domain)*. Both of these present unique difficulties.

Within the classification of laboratories (chapter 3, this volume), GEON can most usefully be understood as a *community infrastructure project*. These are projects whose goals are to develop resources to support scientific work in a particular domain, such as the earth sciences in the case of GEON. Chapter 3 (this volume) identifies three typical organizational difficulties encountered by such endeavors: aligning the research goals of domain scientists and information technologists; determining the best form of management; and producing career rewards and pathways for scientists who help build infrastructure for others to use. Here we are concerned with the first tension—

the tendency of goals to diverge in multidisciplinary teams. We examine the mechanisms that GEON participants have employed to navigate the difficulties noted in chapter 3:

Whose research agenda will be paramount? In partnerships between disciplinary experts and computer scientists, there is often conflict between pursuing the most technologically advanced solutions (which are of research interest to the computer scientists) and more immediately practical solutions.

In addition to the key difficulty identified above—collaboration across computer science/domain boundaries—we also point to the work of collaborating “within” the domain—that is, across the diversity of earth sciences. For many participants, one or both experiences of collaboration are novel. In this section, we characterize the difficulties across each boundary in turn: domain/domain followed by computer science/domain.

#### **Domain/Domain: Toward an Umbrella for the Geosciences**

Scientifically and organizationally, the geosciences span an enormous range of disciplinary configurations. The umbrella term geosciences is deceptively unifying; to say that “geoscientists study the earth,” does not capture the heterogeneity of the natural phenomena and methods that fall under the term. The participants themselves identify over twelve disciplinary specializations within GEON; for instance, these include geophysics, paleobotany, seismology, and petrology. The criteria for knowledge formation and epistemological grounding differs across the geosciences by their history, traditions, and methods of inquiry (Rudwick 1976). Nature does not provide a coordinating framework for science: the methods, language, and concepts of the diverse earth sciences are a matter of culture, learned practice, and social organization (Knorr-Cetina 1999). Below we outline three kinds of disciplinary differences within the earth sciences: social organization, the willingness to share data, and the structure of data.

First, organizationally many of GEON’s earth science participants are located in administrative units of different types. For example, self-identified geoscientists may be housed in geology, physics, or biology departments. This can also mean, for instance, that they have varying degrees of access to computing resources and services, such as whether a research team has information managers or other technical support staff. Such organizational differences make it challenging to build a cyberinfrastructure for the geosciences because the participants do not begin on the same footing for access to technical services or even with a shared familiarity with computing technologies. While a geophysicist may have an entire technical staff, a metamorphic petrologist may never have worked with software more specialized than a spreadsheet. These differences are organizational in that they stand in for the division of labor: Who is responsible for taking care of data? Is taking care of data a dedicated task of a specialized

information manager, or does it compete with the needs of a professor to teach, conduct research, and write articles? Do scientists working for the U.S. Geological Survey have different data practices and research agendas than those based in a university?

Second, scientists have varying traditions for the curation and sharing of their data. A scientist may feel possessive of their data, hoping to draw out future insights, or they may feel uncertain of the quality and thus unwilling to share it with their peers (Borgman 2007; Campbell et al. 2002; Ceci 1988). The extent to which a particular group is prepared to exchange its data varies substantially by discipline. For example, field scientists such as paleobotanists and metamorphic petrologists collect relatively small data sets at particular geographic sites. The intense personal involvement with the research site and the data collection may lead to the unwillingness to contribute such data to a large anonymous repository. They may also feel that the data are incomprehensible or meaningless if not tied to local knowledge about a specific site. On the other hand, instrument-intensive scientists such as geophysicists have established traditions for using large arrays of remote instrumentation, and the discipline has been at the advancing edge of computer science for forty years, from the first analog computers to the first expert system (Bowker 1994). Publicly funded instrumentation often comes with stipulations to release data to the broader community of researchers after a fixed time. Similarly, seismologists have a long tradition of sharing data across both territories and nations: it is in the nature of their data that it does not respect geographic boundaries. Over time, geophysicists and seismologists have developed "cultures" that assume particular data-sharing practices. These varying traditions for data collection, curation, and sharing can seem morally weighted—"the right thing to do"—to the participants. In deciding the policies for an umbrella data repository, at times these varying traditions may even become the object of explicit conflict.

Third, the form and size of databases vary by disciplinary tradition and method. For instance, in mapping topology geologists have begun to use Light Distance and Ranging (LiDAR) scans of the surface of the earth. GEON has developed tools and resources to help geoscientists scale up their data technologies for such approaches. Such techniques generate billions of data points in a "LiDAR point cloud," which today are hosted in an IBM DB2 spatial database running on the DataStar terascale computer at the San Diego Supercomputer Center. In contrast, paleobotanists conduct observations in the field, and collect, classify, and organize samples at a smaller scale. Recently, paleobotanists have used electron microscopy, but most data are not available digitally; thus even "within" paleobotany, data structures will vary broadly. Each method, each disciplinary tradition, and often even each research team will have idiosyncratic methods for transforming the data they collect into databases. Generating tools for working across such diverse data structures is one specific goal of GEON.

Collaboration across the geosciences is an organizational, social, and technical problem requiring an alignment between the practical methods of diverse disciplines, the

institutions in which science is practiced, and the standards that arrange data. To the extent that GEON encompasses heterogeneous earth sciences, the participants have had to articulate and negotiate such differences in building an umbrella for the earth sciences.

#### **Domain/Computer Science: Novel Information Technology for the Earth Sciences**

This brings us to our second axis of collaboration. Building information infrastructure requires domain scientists to work closely with computer scientists. Many of GEON's earth science participants had little or no experience working with computer scientists, and they were unfamiliar with the technical details of information systems or data interoperability.

Computer and earth scientists describe themselves as having different goals, based on reward systems within each research tradition. From the perspective of domain practitioners, computer scientists are disinterested in the practical results of their research or design work. They are said to sit on one side of the "brick wall," designing programs intended for domain use without much consideration for specific application needs, functionality, or accessibility (Suchman 1994). They are able to advance in their own field by publishing their technical innovations in journals, pointing to grants awarded and "demo" programs that stand as surrogates for successful development regardless of practical uptake. Meanwhile, these applications move seamlessly from vaporware to ghostware. Within computer science, the claim goes, little attention is paid to the life cycle of the application in the domain: Has the program been adopted? Does it meet the requirements of the users? Even less consideration is given to providing technical assistance or long-term support for operability (Weedman 1998).

A parallel claim is often set forth about domain scientists: they are rewarded for advances in earth science, but have few incentives to produce and maintain community resources. For example, designing algorithms or visualization packages have generally not been counted toward tenure case decisions. When developing computing resources they instead focus on the development of information technology tools that will serve their particular needs to investigate a scientific research question. What is traditionally rewarded within a scientific community are "science results"—broadly understood as new domain knowledge, or as is frequently stated in GEON, "something new about the Rockies"—rather than the production of long-term information infrastructure.

Within GEON, the problem of reward is often expressed in terms of the future careers of geoscience graduate students participating in the project. These students may have invested much time in creating tools for scientific research, but it is difficult to convey the significance of the contribution to the geoscience community focused on new knowledge. The result may be a graduate student with a record of experience

that is strong within "geoinformatics," but that may appear weak to a traditional geoscience hiring committee.

The aggregate of these two trajectories—information technology's indifference to the domain, and domain scientists' individualist tendencies—amounts to a crucial problem with computer science/domain collaborations. If we want infrastructure to be a long-term, multiuse platform accessible to a community of the targeted users (Bowker and Star 1999; Star and Ruhleder 1994), the computer scientist must be configured to care about science implementation success as measured from *within* the domain, and each scientist must be motivated to care about creating infrastructure resources for a broader scientific community. Building technology that is usable in practice must matter to a computer scientist, and designing technology for a broader community must matter to an earth scientist.

Across the domain/domain and computer science/domain divides, GEON participants have had the task of creating an umbrella infrastructure for the earth sciences. Each of the next three sections identifies phases in the development of GEON, and how in each phase the participants worked to cross both boundaries. We begin before GEON, as the participants articulated a vision of collaboration in the proposal. We then outline the initial meetings, as the members sought to find means for communication across disciplinary boundaries. Finally, we explore the work around GEON's test beds, which helped to coordinate activity across both sets of boundaries.

#### Making a Vision: Writing the GEON Proposal

In this section, we trace the work of GEON's PIs as they sought to create a vision of multidisciplinary collaboration that balanced computer and domain science research. As the twelve PIs of the grant wrote the proposal, they were continuously aware of a *community opinion* that placed doubts on GEON as a *contribution* to the computer or earth sciences. In many senses, this was a formative controversy that occurred prior to the funding of GEON. The controversy has shaped the goals and methods of the project (Collins 1981; Scott, Richards, and Martin 1990). As in other fields, domain scientists have often felt—whatever assurances were given to the contrary—that money spent on computing resources was money not being spent on science. The balance between computer science and domain remains an ongoing concern for the GEON participants, but this concern was first articulated in the multiple iterations of writing the grant proposal.

The controversy was drawn along and across disciplinary lines: Is research in GEON geoscience? An exercise in computer science? Both? Or perhaps neither? The lines of debate can be summarized into two prevalent disciplinary arguments:

- GEON is not engaged in computer science research but merely in the application of information technology to geoscience problems and research

- GEON is not engaged in geoscience research but in experimenting with information technologies not yet sufficiently developed to contribute to practical earth science questions

To understand how this debate emerged, we must first turn to the history of the funding program for GEON.

Beginning from early planning meetings in 2000, the GEON PIs decided that the vision of the proposal would be to place the research goals of computer and domain science on equal footing. This goal was in marked contrast to more traditional "science-centered" or "technology-driven" projects. Computer scientists would not be "mere technicians," and geoscientists would not be sites for testing novel IT. In the practice of composing the proposal, the PIs found the task of satisfying both groups more daunting than initially envisioned; yet doing so was encouraged by the structure of the funding itself.

GEON was funded under the NSF's Information Technology Research (ITR) program, in which basic research remained a central goal. The requests for proposals issued as part of the ITR program specifically demanded new, experimental, and high-risk research. In order to justify GEON as an ITR project, the proposal writers had to demonstrate that the project would address important geoscience questions in addition to those of producing infrastructure. Because a part of its funds would come from computer science and the other part from the earth sciences, GEON would also have a double responsibility. As noted by one of the PIs, GEON would have to satisfy two sets of scientific criteria: "The RFP [request for proposals] from ITR was very clear, they want something risky, experimental, from the IT side, but we wanted GEO to fund us, too, and that meant they had to feel like we were doing something about geology, or for the earth science community."

GEON PIs described the difficulty as a tension between research and development (Lawrence 2006). Scientific infrastructure is meant to offer a relatively stable and transparent base for research. If GEON was to be a platform for geoscience, it would have to be accessible to the "average earth scientist"—supporting everyday work and making data accessible in a straightforward manner. Yet would such a stable set of technologies meet the criteria of computer science research that focuses on novel capacities? On the other hand, if the proposal placed too great an emphasis on the contemporary research questions of computer science, geoscientists would see it as experimental rather than stable.

In such a scenario, what would a balance between computer science and domain research look like? It was not possible to answer this question in advance; rather, it was carefully crafted and negotiated over two iterations of the proposal-writing process.

The first GEON proposal was explicitly declined by NSF on the grounds of a poor balance between computer science/domain research. One of the GEON PIs stated that reviewers noted that the first proposal leaned too far in the direction of computer



science research: "We threw in a lot of computer science and digital libraries kinds of research, and we also threw in a whole lot of earth science. We thought we were covering our bases, but they [the review panel] thought we were stretching ourselves thin."<sup>2</sup> Why should the NSF Directorate for Geoscience fund information technology research when its mandate is clearly to generate new knowledge about the earth? Within computer science a similar, if muted, debate played out over whether the infrastructure building was a contribution to knowledge of computing. If GEON relied on relatively established hardware and software technologies, then it came closer to an implementation project rather than computer science research. In this case, why should the Directorate for Computer and Information Science and Engineering fund this project?

The second ITR proposal had to demonstrate stronger geoscience contributions without going too far in the other direction and causing disinterest from its computer science constituency. Participants' most noteworthy response to the first proposal's rejection was to tie information technology research to specific sites of geological science; these came to be known as the mid-Atlantic and the Rocky mountain *test beds*. The test beds were chosen because they were important sites for geoscientific research and had been studied for over a century. Further, each test bed had well-documented science questions drawn from the various disciplines making up GEON (domain/domain balance), and had multiple data sets in each area. For example, data integration and visualization in the mid-Atlantic can serve to help bring together data, evidence, and scientists in current controversies in seismology involving continental growth through accretion: "To assess this mechanism of continental growth, the models must be constrained by multidisciplinary observational data sets within the region that will be assembled by the GEON geoscience team" (Keller 2003, n.p.). Here we can see the identification of a problem set, the communities that are participating in a theoretical debate, and at least a suggestion of the formal models and data sets that would be required for technical integration. The test beds could help to bring all three elements together.

The test beds were a highly successful addition to the second GEON proposal. The PIs have noted that the reviewers responded well to them. More important for the GEON participants, the test beds served to "balance" the vision of collaboration for the project, giving the computer scientists specific data sets to work with, and the earth scientists particular geographic regions to focus their research.

Through the negotiation process, GEON became imbued with this particular vision of the computer science/domain collaborative construction of infrastructure. This vision was articulated in its grant proposal through the allotment of funds, the distribution of responsibility, and the promises for "deliverables" in the form of computer and earth sciences knowledge. Such a vision was carefully articulated among participants. Thus, more than simply serving to secure funding for the project, the proposal writing was a constitutive process for participants. Over the years, members have regularly returned to the proposal as a resource for articulating the GEON collaborative vision.

### Proposal to Collaboration: From a Shared Vision to a Community of Practice

Above we described the crafting of a shared vision of multidisciplinary collaboration through proposal writing. Still, a proposed *plan* on paper was not the *activity* of working together. While having the proposal funded was the formal inception, the next step was moving from a paper GEON to a practical organization. As with writing the proposal, the practice of multidisciplinary collaboration was an accomplishment. In this section we focus on the *meetings of GEON* as the sites for organizing from a funded proposal to a practical collaboration. Such gatherings served to foster the shared understandings that are the keystones of multidisciplinary collaboration; Jean Lave and Etienne Wenger (Lave and Wenger 1991) draw attention to mutual learning in such groups and call these communities of practice.

Lave and Wenger (1991) have discussed communities of practice as a useful unit of analysis for understanding how heterogeneous groups come together to perform complex tasks. In particular, they focus on the process of social learning that occurs when people who have a common interest in some subject or problem collaborate over an extended period to share ideas, find solutions, and build innovations. Contrary to received understandings, community should not be considered a common starting point for organized action but the reverse: community is the *product* of shared activity, collectively negotiated meanings and purposes, and trajectories of shared learning. Above we have already described the *differences* in the research methods and goals of the GEON participants, primarily characterized by the boundaries computer science/domain and domain/domain; in practical collaborative work, the concept of a community of practice emphasizes how members come to work across those boundaries.

The first GEON meeting was the kickoff, and it was structured to help the participants begin communicating across their disciplinary specializations. Through such gatherings and over time, the participants articulated an understanding of the collective challenges facing GEON. While the PIs had collaborated in writing the proposal, the kickoff meeting had more ambitious goals: it brought together the research teams of the PIs, and provided an opportunity for individual participants to communicate their research questions and collectively debate the understanding of "what GEON is to be."

The kickoff meeting served to introduce the information technology team and its planned technologies to the geoscientists. In turn, the geoscientists presented their science questions and some initial descriptions of the kinds of data necessary to achieve their goals (see figure 17.1).

The great majority of the time at this meeting was spent as individual scientists (earth and computer alike) presented their research to the larger team. So, for example, a geoscientist described the mechanisms for pluton formation—an igneous rock body formed at great depth and intruding among others. A computer scientist described

# Problem: Scientific data integration from Questions to Queries

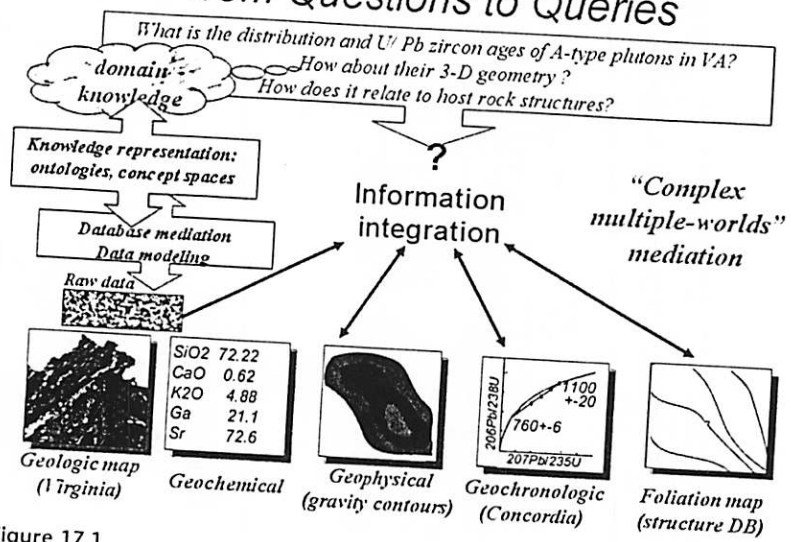


Figure 17.1 Elaborate slide shows are composed to share technical knowledge with the heterogeneous experts of GEON; this slide captures the notion of data integration

ontologies—an information technology for representing concepts as semantic relations. Each presentation was followed by a discussion period, allowing the participants to ask for clarification or draw links across their research. Along with computer science and geoscience introductions, this meeting also served to begin coordinating research goals among earth scientists: Would GEON’s lithology ontologies draw from the already-established British rock classification? How could LiDAR data-cleaning techniques be used for applications mapping the topology of drainage basins? The participants discussed the diverse research conducted by GEON earth scientists and considered the ways in which individual research goals could be aligned with those of the larger project.

In a received understanding of community, members coming from similar formal positions and backgrounds already hold common beliefs. In contrast, the notion of a community of practice emphasizes the *production* of a community identity through shared learning experiences and explicit discussions of common purpose. From this perspective, communities are not so much about extant homogeneity as about generating a common orientation for members. A key feature of the kickoff meetings was the inclusion of a broader array of participants from the geosciences than the PIs. A community infrastructure involves a wider set of participants than those we have identified

as "computer science and domain." Such activities always draw on technical and administrative resources that are often later forgotten in accounts of scientific and technological activity, but that are seminal in building successful infrastructure.

The kickoff meeting and subsequent annual meetings encompassed graduate students, information managers, ancillary research staff, and the project administrators. We can only identify these as GEON community members after the fact. They began from highly heterogeneous starting points: scientists/professors, graduate students, technicians, and information managers—all of whom had to be successfully integrated into building up and maintaining a functioning cyberinfrastructure. As was typical for such collaborations, the set of personnel was ever shifting—and the boundaries between participating, lurking, and ignoring were fluid. For example, information managers are the curators of geoscience data, specifically tasked with the responsibility of preserving databases and helping with access. The diversity of speakers at GEON meetings assisted in bringing such "long-term" concerns to the table in matters of planning and design. Including heterogeneous participants during the process of design means they can stand-in for the diversity of downstream users in GEON's future as a functioning infrastructure. Through the learning environment created in the initial meetings, these heterogeneous participants *became* a community despite their initial differences.

From a narrow instrumentalist perspective of "doing research" or even "building infrastructure," we could say that "nothing gets done" in such meetings; however, presentations on scientific and technical specializations came to shape a common foundation for the heterogeneous participants. They built up common knowledge and vocabularies for the multiple participating disciplines. As the tasks shifted from introductions to practical collaborations, the participants were able to begin from a set of basic understandings about the research questions constituting GEON's base of expertise. This kind of exchange proved doubly valuable in the long term for future relations between computer science and domain. Specifically, computer scientists came to have a much deeper understanding of how the domain functions, the kinds of work that geoscientists do, and the way they produce knowledge.

#### **Collaboration to Work: Coordinating Activity across Disciplinary Boundaries**

We have covered the development of a shared vision for collaboration and the emergence of a community of practice. What remains for multidisciplinary activity is the organization of the work itself. At the core of GEON is the production of novel technological resources for the domain in the form of information tools that will support geoscientific work.

A key goal in the GEON project is achieving interoperability. Interoperating the geosciences means producing computer-mediated linkages between the multiplicity of data sets and the computing resources in the earth sciences. This has included, for

example, creating schemas that permit queries for simultaneously searching multiple databases. By making heterogeneous data sets interoperable two prospects emerge: transforming everyday work in the geoscience fields to answer questions that cannot be approached by one discipline alone, and acting as proof that information technology can be used to facilitate interdisciplinary synthesis. In this section we ask: How are such technologies collaboratively designed while meeting the disciplinary requirements of earth scientists (new knowledge about the earth) and computer scientists (new knowledge about computation)? There are, of course, multiple approaches adopted by GEON participants; here we return to the test beds in order to examine how these served to coordinate activity across disciplinary boundaries.

To elaborate this example we draw on the concept of "boundary objects." Susan Leigh Star and James Greisemer (1989) described boundary objects as sustaining multiple meanings for heterogeneous members while also meeting the informational requirements of all. The notion helps to explain the collaboration of expert groups, such as scientists, without relying on the assumption that they need to establish a consensus of purpose or technical meaning among themselves. This is a good way to understand how the test beds serve to coordinate work in GEON across two sets of disciplinary boundaries: computer science/domain and domain/domain. For earth scientists the test beds were sites of empirical research in many subfields, while for computer scientists the test beds collected data sets in multiple formats; in sustaining different meanings for these participants, the test beds also coordinated activity *within* the geosciences and *across* to computer science.

As noted, the test beds are two geographic regions in the United States: the mid-Atlantic and the Rocky Mountains (see figure 17.2). They were added to GEON during the second round of proposal writing with the specific intention of grounding applications of information technology to sites of empirical geoscience research. In the earth science the test beds are sites of investigation for many subfields, and thus they meet the requirements for disciplinary diversity. For instance, the Rocky Mountains are a prime site for orogenic processes—or mountain building—between the stable interior of North America and the plate margin of the West Coast; it is also the site of multiple fossil deposits; and a rich stratigraphic record has been captured, over the millennia, as intraplate stresses have superimposed basins. Geologically, then, the Rocky Mountain test bed is a site of contemporary and historical inquiry for seismologists, geophysicists, and paleobotanists alike. While varying in their specific interests and methods, many earth scientists could come together to conduct their research within this test bed: the Rocky Mountain was an umbrella (domain/domain) for various empirical concerns.

Conversely, for the database specialists of GEON, the Rocky Mountain test bed provided multiple complex data sets. In their decades of research, seismologists, geophysicists, and paleobotanists have built up collections of data about the area; for them it

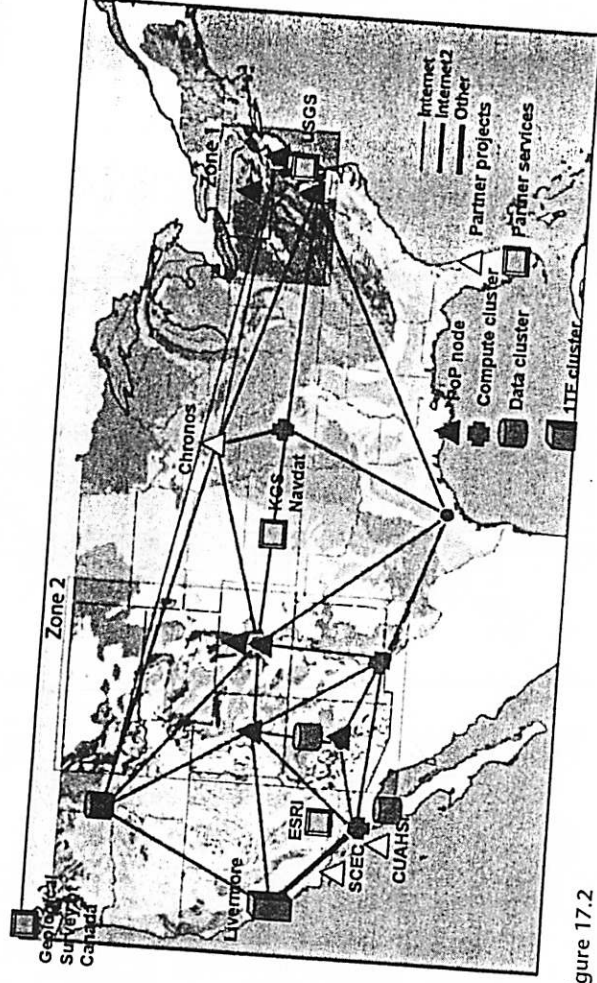


Figure 17.2

GEON partners and the two test beds: Rocky Mountain on the left, and mid-Atlantic on the right

was the geologic formations that were of interest, while for GEON's computer scientists it was the databases themselves. The varying schemas and semantics of these data collections offered compelling research scenarios for the information technology team. It challenged them to devise approaches to create semantic and logical interoperability across heterogeneous databases. A GEON user can search multiple integrated data formats, including shapefiles, relational databases, WMS maps, Excel files, geoTIFFs, ASCII files, and netCDF files. In addition, users can search across multiple databases. For example, GEON hosts the PaleoIntegration project, which enables interoperability among several global fossil and sedimentary rock data sets (Rees 2007).

The test beds provided organizational cues for producing an alignment between particular scientific questions and information technologies. Through them, information technology applications were driven by geoscience questions; in GEON they called this the "science driver model." Computer science developments became grounded through science drivers and the specificities of domain information structures rather than abstracted understandings of an imagined geoscience. For example, in the kickoff meeting described earlier, earth scientists dedicated an afternoon to discussing the data sets that would be required for addressing the question, "What is the nature of interactions among Paleozoic, Laramide, and late Cenozoic basins?" The participants identified six distinct databases that the information technologists could use to focus their integration work. This first step of outlining science questions along with their corre-

sponding databases was followed by iterative discussions as the geoscientists came to specify their questions in greater detail and the information technologists came to understand the databases at a finer granularity.

Thus, test beds and the science driver model of development, served as devices for managing the distribution of resources. As we have noted, GEON's computer scientists are high-end researchers rather than technical support workers. Yet even though information technology was highlighted in the GEON proposal, there was also an acknowledgment that in building infrastructure, low-level technical support work would also be necessary. Much of the basic work of integrating databases, for instance, was not cutting-edge research for computer scientists. In other words, this work was considered technical support rather than the generation of new, publishable knowledge. Working to integrate data from the test beds afforded compelling research questions, but it also demanded a certain amount of basic implementation.

The GEON proposal necessarily acknowledged some low-level service provision by the information technology team: "While not viewed as IT research, format and schema incompatibility issues will be addressed specifically for the test bed-related databases." (GEON, 2002, 5). As described above, for information technology participants, there was a thin but important line between providing the support necessary for the research and slipping into a service capacity. Domain scientists who have extensive experience working with information technology, including sophisticated technology for data collection or the management of databases, often consider information technology as a technological substratum to be delegated to a specialized staff. GEON's philosophy, however, was of an equal partnership between computer and earth scientists.

With the increasing involvement and complexity of information systems in scientific work this relationship is changing, and in community infrastructure-building projects, the relationship must necessarily be reconfigured. But old habits die hard. The particular work and content of these technicians often remains invisible to scientists (Shapin 1989); scientists may pass on "to do" lists for the technicians rather than engage in a close collaboration. On many occasions in the early period of GEON's development, the geoscientists requested particular "low-end" services from the information technology team such as building Web sites. This resulted in high-priced and high-end computer science researchers who specialize in, for instance, semantic mapping, creating low-end front-ends for Web service databases. The outcome was unsatisfying to all the participants: computer scientists had no interest in writing Web pages and so produced mediocre outputs, while geoscientists became dissatisfied with the technical team's inability to provide quality service.

The problem was of poorly aligned expectations. The management of information technology resources for GEON was the site of several skirmishes around the understanding of the nature of the project. For some of the earth science GEON members,

the information technology team appeared as a source for bountiful high-tech service, such as Web pages, user interfaces, and data set cleanup. Many of these requests by the geoscientists were communicated directly to the information technology team of GEON, rather than through the information technology PIs. As the larger goals and organizational vision of GEON solidified, it became clear that such projects fell outside the bounds of GEON's work. And yet how does one withhold information technology resources from a participant's work without causing strife? This resulted in tensions about the role of the information technology team within GEON, perceptions of efficacy, and the reopening of further debates about GEON's mandate.

Once again a return to the GEON vision of dual computer and geoscience research goals was necessary. In order to do so, structures and routines to manage the distribution of information technology were established. In this case an organizational "protective boundary" was built around the information technology resources. In the second and third years of GEON (2003 and 2004), as specific science goals solidified around the two original test beds, these became focal points for resource distribution. Decisions about what did or did not count as a relevant test bed development of information technology resources were delegated to two earth science PIs. Requests for any development efforts on the part of the geosciences participants went through one of the two test bed representatives, who in turn made executive decisions about the importance of the request relative to the timeline, the significance to the project, and the labor required, and then passed on those requests deemed necessary to the information technology team. In order to gain traction, visions of science-driven information technology development had to be grounded in practical arrangements.

The test beds sustained scientific meanings for all the participants: they are comprised of rich fossil sites for paleobotany; they are complex stratigraphic formations for seismology; and they are data integration problems for computer science. By simultaneously meeting the research needs of diverse groups, the test beds can be described as boundary objects, facilitating coordination across disciplinary interests without necessitating a "translation" or erasing of difference. As Star and Griesemer (1989, 388) observed in their studies of multidisciplinary collaboration, "We find that scientific work neither loses its internal diversity nor is consequently retarded by lack of consensus." They describe cooperative work in such settings as operating in an "ecology," by which they mean that one group's views need not necessarily be devalued relative to the other: it is not that computer science is in the service of geoscience but rather that both parties seek to address contemporary research problems. The concept of boundary objects allows us to understand the coordination of heterogeneous scientific work in such a manner as to serve multiple sustained interests. As we have seen, boundary objects do not "foreclose conflict" but rather provide practical and material points of shared orientation, affording mechanisms for the navigation of disagreement.



## Conclusion

It is paradoxical that most cyberinfrastructure projects have relatively short time horizons, while all the evidence is that the work of building infrastructure is inevitably long and complex. It is an open question how long it takes genuinely new questions to emerge from a new information infrastructure. According to previous research on the productivity paradox in computing (i.e., productivity goes down when computers are introduced into the workplace and may require decades to reverse this trend), we might expect changes to require at least fifteen to twenty years (David 1989). Projects such as GEON, however, mark a clearly innovative trajectory that recognizes the centrality of community-building work; the PIs have evolved some of the organizational strategies that will be necessary to achieve this end as well as move beyond narrow conceptions of "tubes and wires" infrastructure. The next phase of GEON, sometimes called GEON2, aims to populate the databases, polish the interfaces, and people the community of researchers.

The work of community building is time-consuming in itself. In writing the proposal a vision was set, forming a basis for a community of practice; this took months of work. In ambitious multidisciplinary endeavors, such as infrastructure building, our findings suggest that time and resources must be appropriately allocated to carefully plan crossings of domain/domain and computer science/domain boundaries.

In this chapter, we have traced three moments in fostering such multidisciplinary collaborations. This process began before the project through proposal writing. We have argued that proposal writing was a form of collaborative work that established the criteria, organization, and distribution of labor for future activity in the project. Over the years of GEON, the participants regularly came back to the proposal in moments of identity crisis. Whenever it appeared that efforts were leaning too heavily in either the direction of earth or computer science, the proposal served as a reference point for a computer science/domain balance by which to return focus to "learning something new about the Rockies" or "doing some serious computer science."

Moving from the vision articulated in the proposal to the action and organization of a large-scale infrastructure project was the next step. The GEON meetings offered a venue as the participants within geology and across to computer science shared their research agendas and topics. It is difficult to measure or formally evaluate progress achieved in such early meetings. We contend that the shared learning experiences in meetings, of demonstrating diverse research agendas, topics, and methods, later proved invaluable to long-term multidisciplinary collaborations. Efforts to make accessible and relatively transparent the technical specializations—*among* domain scientists and *with* computer scientists—were seminal to the more practical work of building infrastructure.

Building scientific infrastructure involves a kind of collaboration that must foster deep understanding. This is as true across domain and computer science as it is within the domains. It is the information technologies themselves that presented many of the unforeseen complications in building GEON. While within the theoretical computer science literature many of the base technologies of GEON were considered resolved (see for example Sheth 1999), there remained a significant gap between theoretical resolution and enacting production quality systems (Fountain 2001). These technologies simultaneously presented a source of complication and communicative difficulty between information technology and domain, and occasions for aligning understandings through discourse. In projects where there are multiple competing interests, as in the case of GEON, which is earth and computer science research, sites such as the test beds serve in the common coordination of work. We have labeled such coordination sites boundary objects for their ability to simultaneously preserve multiple meetings and satisfy the informational requirements of competing interest groups. It is often assumed that collaboration requires consensus; the concept of boundary objects and the processual nature of a community of practice in GEON reminds us that this is not necessarily the case. Rather, collaboration across disciplinary difference can be a matter of generating shared orientation (or community identity) through common tasks or projects.

#### Acknowledgments

We would like to thank Karen Baker, Chaitan Baru, Steve Epstein, Florence Millerand, and the editors of this volume for their comments on drafts of this chapter. This material is based on work supported by the NSF under grants #04-33369 ("Interoperability Strategies for Scientific Cyberinfrastructure: A Comparative Study"), #0525985 ("The Standardized Revolution of Science: Building Cyberinfrastructure for the Geosciences"), and #0225673 ("GEON"). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors, and do not necessarily reflect the views of the NSF.

#### Notes

1. The term domain is a coinage from within computer science and information technology circles to refer, somewhat generically, to fields of application—in this case, the earth sciences. We adopt this actor's category with some ambivalence, as on occasion earth scientists themselves have chafed at the term. Additionally, we use the terms earth science and geoscience interchangeably, as do the participants, to refer to the various fields that GEON seeks to bring under its umbrella infrastructure.
2. Actual proposal reviews are considered confidential material at the NSF. The data on "why" the first proposal was not funded is based on secondhand accounts.

## References

- Borgman, C. L. 2007. *Scholarship in the digital age: Information, infrastructure, and the Internet*. Cambridge, MA: MIT Press.
- Bowker, G. C. 1994. *Science on the run: Information management and industrial geophysics at Schlumberger, 1920–1940*. Cambridge, MA: MIT Press.
- Bowker, G. C. 2000. Biodiversity datadiversity. *Social Studies of Science* 30 (5): 643–683.
- Bowker, G. C., and S. L. Star. 1999. *Sorting things out: Classification and its consequences*. Cambridge, MA: MIT Press.
- Campbell, E. G., B. R. Clarridge, M. Gokhale, L. Birenbaum, S. Hilgartner, N. A. Holtzman et al. 2002. Data withholding in academic genetics: Evidence from a national survey. *JAMA* 287 (4): 473–480.
- Ceci, S. 1988. Scientists' attitudes toward data sharing. *Science, Technology, and Human Values* 13 (1–2): 45–52.
- Clarke, A. 2005. *Situational analysis: Grounded theory after the postmodern turn*. Thousand Oaks, CA: Sage Publications.
- Collins, H. M., ed. 1981. Knowledge and controversy: Studies of modern natural science. Special issue, *Social Studies of Science* 11 (1).
- David, P. A. 1989. Computer and dynamo: The modern productivity paradox in a not-too-distant mirror. Discussion paper series no. 172. Stanford, CA: Stanford University, Center for Economic Policy Research.
- Fountain, J. E. 2001. *Building the virtual state: Information technology and institutional change*. Washington, DC: Brookings Institution Press.
- Keller, G. R. 2003. GEON (GEOScience Network): A first step in creating cyberinfrastructure for the geosciences. *Electronic Seismologist* (July–August). Available at <[http://www.seismosoc.org/publications/SRL/SRL\\_74/srl\\_74-4\\_es.html](http://www.seismosoc.org/publications/SRL/SRL_74/srl_74-4_es.html)> (accessed June 19, 2007).
- Knorr Cetina, K. 1999. *Epistemic cultures: How the sciences make knowledge*. Cambridge, MA: Harvard University Press.
- Lave, J., and E. Wenger. 1991. *Situated learning: Legitimate peripheral participation*. New York: Cambridge University Press.
- Lawrence, K. A. 2006. Walking the tightrope: The balancing acts of a large e-research project. *Computer Supported Cooperative Work* 15 (4): 385–411.
- Rees, A. 2007. Phanerozoic earth and life: The PaleoIntegration Project. Paper presented at geoinformatics 2007 conference, San Diego, California, May. Available at <[http://gsa.confex.com/gsa/2007GE/finalprogram/abstract\\_121912.htm](http://gsa.confex.com/gsa/2007GE/finalprogram/abstract_121912.htm)> (accessed June 19, 2007).
- Ribes, D., and K. S. Baker. 2007. Modes of social science engagement in community infrastructure design. In *Proceedings of the third communities and technology conference*, ed. C. Steinfield, B. Pentland, M. Ackerman, and N. Contractor, 107–130. New York: Springer.

- Ribes, D., K. S. Baker, F. Miller, and G. C. Bowker. 2005. Comparative Interoperability Project: Configurations of community, technology, organization. In *Proceedings of the fifth ACM/IEEE-CS joint conference on digital libraries*, 65–66. New York: ACM Press.
- Ribes, D., and T. A. Finholt. 2007. Tensions across the scales: Planning infrastructure for the long term. In *Proceedings of the 2007 International ACM SIGGROUP Conference on Supporting Group Work*, 229–238. New York: ACM Press.
- Rudwick, M. 1976. The emergence of a visual language for geological science, 1760–1840. *History of Science* 14:149–195.
- Scott, P., E. Richards, and B. Martin. 1990. Captives of controversy: The myth of the neutral social researcher in contemporary scientific controversies. *Science, Technology, and Human Values* 15:474–494.
- Shapin, S. 1989. The invisible technician. *American Scientist* 77:554–563.
- Sheth, A. P. 1999. Changing focus on interoperability in information systems: From system, syntax, structure to semantics. In *Interoperating geographic information systems*, ed. M. F. Goodchild, M. J. Egenhofer, R. Fegeas, and C. Kottman, 165–180. Boston: Kluwer Academic Publishers.
- Star, S. L. 1999. The ethnography of infrastructure. *American Behavioral Scientist* 43:377–391.
- Star, S. L., and J. R. Griesemer. 1989. Institutional ecology, “translations,” and boundary objects: Amateurs and professionals in Berkeley’s Museum of Vertebrate Zoology, 1907–39. *Social Studies of Science* 19:387–420.
- Star, S. L., and K. Ruhleder. 1994. Steps toward an ecology of infrastructure: Complex problems in design and access for large-scale collaborative systems. In *Proceedings CSCW 1994*, 253–264. New York: ACM Press.
- Suchman, L. 1994. Working relations of technology production and use. *Computer Supported Cooperative Work* 2 (1–2): 21–39.
- Weedman, J. 1998. The structure of incentive: Design and client roles in application-oriented research. *Science, Technology, and Human Values* 23 (3): 315–345.

# **Scientific Collaboration on the Internet**

**edited by Gary M. Olson, Ann Zimmerman, and Nathan Bos**

**The MIT Press  
Cambridge, Massachusetts  
London, England**