Tensions across the Scales: Planning Infrastructure for the Long-Term

David Ribes
School of Information
University of Michigan
1075 Beal Avenue
Ann Arbor, MI 48109-2112, USA
1.734.764.1383
dribes@umich.edu

Thomas A. Finholt
School of Information
University of Michigan
1075 Beal Avenue
Ann Arbor, MI 48109-2112, USA
1.734.764.6131
finholt@umich.edu

ABSTRACT
In designing information infrastructure participants are planning for the long-term. The notion of infrastructure evokes images beyond 'a proof of concept,' a 'prototype' or an isolated 'application'; it is intended to be a persistent, ubiquitous and reliable environment. However, in implementing such projects participants confront multiple difficulties such as securing sustained funding, supporting maintenance and integrating new technologies.

Based on cross-case ethnographic analysis this paper traces nine tensions identified by participants as they endeavor to transition from short-term projects to long-term information infrastructure. We explore three core concerns framed by actors: motivating contribution; aligning end-goals; and designing for use. These concerns have unique implications for each scale of infrastructure: institutionalization; the organization of work; and enacting technology.

Categories and Subject Descriptors
K.4.3 [Training] Organizational Impacts *Computer-supported collaborative work*

General Terms
Management, Design, Human Factors, Theory

Keywords
Long-term, Sustainability, Cyberinfrastructure, Science, eScience, Communities, Ethnography, Cross-case analysis

1. INTRODUCTION
Under the rubric of cyberinfrastructure (CI) in the US and e-Science in the UK there is currently a significant emphasis placed on the development of long-term information infrastructure for the support of scientific, academic and scholarly research. These endeavors are centered on providing computing resources, facilitating long distance collaboration, enabling data sharing, and developing novel analytic tools.

CI projects often make claims of 'revolutionizing research,' engendering a 'paradigm shift' or transforming scientific practice[1], however, many of these projects remain in early stages of development. The resources and work that will be necessary for successful implementation and adoption of CI, then, are still unclear. Because efforts to systematically develop cyberinfrastructure for the sciences are relatively recent, participants find themselves struggling to identify and then articulate their difficulties.

An emerging body of social and organizational research is focusing on these problems [2, 3], but the field remains sparsely researched. In this paper we focus on the tensions experienced by participants in cyberinfrastructure development projects as they seek to transition from short-term projects to long-term facilities. We explore three primary *concerns* of actors and offer a tripartite set of sensitizing concepts – the scales of infrastructure – for the analysis of cyberinfrastructure development. At the intersection of the concerns and scales we find nine *tensions* in developing long-term information infrastructure. Table 1 summarizes the tensions.Our data are drawn from *in situ* observation of the work of designers, project managers, and other participants in cyberinfrastructure design, development, deployment and use. Our particular focus is the *work of participants* as they plan and design infrastructure to last for the long-term. Our research is comparative – we draw our insights from studies of multiple ongoing projects – and has ethnographic goals: by capturing the orientation of participants we seek to convey the tensions experienced in cyberinfrastructure design. These tensions are not hidden; rather, they are regularly discussed by participants but only rarely addressed in scholarly writings. Finally, because studies of cyberinfrastructure are nascent we cannot yet know how to successfully plan and implement systems meant to be used for decades. Therefore, rather than offering a programmatic response ('how to design for the long-term' or 'factors in CI success') we first seek to frame the difficulties actors encounter on the ground and thus begin the process of defining researchable questions around CI development.

2. THE LONG-TERM ACROSS THE SCALES OF INFRASTRUCTURE

Infrastructure requires both a more specific and also more expansive conceptualization than 'technology.' Infrastructure stretches across multiple scales of action: i- it is a technological venture, seeking to deploy durable resources to support work, automate tedious tasks and enable collaboration; ii- it is a matter of human work, organization and maintenance, or as sociologist Susan Leigh Star reminds us, one person's infrastructure is another person's daily routine of upkeep [4]; and, iii- it is an institutional venture, seeking to provide stable and accessible services to communities at national and international levels. Environmentalist Stewart Brand mirrors our distinction between infrastructure and technology in his description of an iconic clock intended to last ten thousand years. In characterizing the "millennium clock" Brand observes "We're going back to the early excitement about clocks - they were big, they were monumental, they were something that a city would organize itself around - the clock in Prague, the clock in Venice" [5]. In this analysis Brand takes into account the technology of the clock itself but he stretches his frame to include the community and social organization; the latter two sustain the operation and meaning of 'an early clock.' Similarly, to render infrastructure researchable we must broaden our analytic gaze across the scales of action – institutionalizing, organizing work and enacting technology – and explore temporal dimensions, particularly 'the long-term' [6].

Enacting Technology – In terms of funding and expertise, the technological aspects of information infrastructure have received the greatest attention. For example, proposals for CI often include extensive discussions of data management and exchange, e.g., how to develop tools to effectively share data across generations of participants [7], or the scalability problem, e.g., how to ensure that systems can accommodate the demands of increased numbers of users [8]. Data integration, metadata and ontologies [9] help to address the question of how we will link the heterogeneous databases of the sciences, how we will preserve legacy archives, and how we will communicate across disciplinary boundaries. Research in these areas has produced progress for the persistence of technology, capturing organizational memory [10], planning for growth, and more generally, thinking beyond months or years. However, the common hubris surrounding these technical solutions often hides complications experienced by developers and users of CI. Following Jane Fountain's research on implementation efforts in digital government we describe the practical activity of developing stable, usable infrastructure as *enactment* [11]. The notion of enactment draws attention to the work that is necessary to shift from experimental technologies to a functioning, stable infrastructure available for everyday use.

Organizing Work – Infrastructure development (whether research, design or deployment) is a matter of practical work. Its accomplishment is the ordinary daily activity of participants [12]. Thus actors, particularly those in managerial roles, are regularly concerned with how the internal organizational arrangements serve to motivate participants and produce outcomes consistent with current developmental goals. Lougher and Rodden note "The success of most projects is often dependent on the ability to effectively organize the activities of the specialists involved. As projects grow in size, complexity and life-span this is becoming an increasingly difficult task" [13:228].

In thinking of persistent infrastructure we must expand beyond the 'tubes and wires' of technology to its organizational setting. Just as Brand's clock does not stand apart from its community ('the clock in Venice'), a physical infrastructure is enmeshed with the routines and practical work of its use, upkeep and repair. Studies of computer-supported cooperative work (CSCW) have primarily approached the matter of the long-term as a question at the intersection of technology and practice. For example, the long-term is often framed as a knowledge management problem: how to develop tools which effectively serve to share information across generations of participants [13], how to preserve and communicate design rationale as individual new members join and old members retire [14]. However such approaches narrow the frame of analysis: 'knowledge management' becomes reduced to the tools by which it is conducted. In contrast, in this paper we focus on the *kinds of work* in which participants engage: conducting novel scientific research, developing applications or 'hardening' and maintaining systems[1]. For instance, each of the four infrastructure projects we explore in this paper has a knowledge management component: technologically based solutions to support long-term collaboration. However, in these projects knowledge management technologies are coupled with a concern for sustaining them over time, what Bowker has called *memory practices* [18]: organizational arrangements, techniques and routines for preserving and maintaining an accessible archive, repository or database.

Institutionalizing – Along with the technical and practical components, we must further expand the frame to include the goal of these projects to achieve persistent institutional arrangements: these are customs and regularized behaviors important to a collective [19] and organizations linked to the state which perform a public service [20]. The goal of CI is to institutionalize informational resources for the sciences. Institutionalization is the work of generating sustainable goods and services linked to social or collective purposes and with connotations of permanence, transcending individual lives, interests or intentions [21]. Thus, in our cases scientific information infrastructure has institutional goals [22]: it is intended to generate public goods through the support of research; it should endure beyond any particular scientific question; and, is linked to governance and sustained state funding. Working to achieve institutional goals is not only a 'top-down' matter for those in policy spheres, it also a practical, daily concern for participants in design [23]. Participants in infrastructure design projects are regularly engaged in securing funding, marshalling support from domain communities[2] and

---

[1] There have been various commonly identified difficulties within infrastructure development projects. Because many contemporary CI projects are geographically distributed and composed of participants with heterogenous expertise, much of the research on work in CI development has been around the issues of communication, collaboration, coordination [15-17]. In this paper distanced collaboration is discussed only to the extent that it relates to long-term sustainability.

[2] Within CI circles 'the domain' refers generically to the community to be served by IT: in our cases scientists and engineers. In this paper we adopt this actor's category despite occasional chafing on the part of those labeled by the term.

Table 1: We have parsed the nine tensions identified by actors in long-term development by their persistent concerns and the scales of infrastructure.

Concerns / Scales	Institutionalizing	Organizing Work	Enacting Technology
Aligning End-Goals	Project vs. facility	Planned vs. emergent	Inclusion vs. readiness
Motivating Contribution	Individual vs. community	Development vs. maintenance	Research vs. production quality systems
Designing for Use	Communities vs. constituencies	Research vs. development	Today's requirements vs. tomorrow's users

actively seeking to transform the norms and customs of their colleagues.

***

The scales of infrastructure reflect the range of participants' activity as they go about the work of research, design and deployment. However, we do not argue that the scales operate through independent logics; in planning, designing and implementing CI actors traverse and link across the scales [24]. The scales, then, are *sensitizing concepts* [25] generated from observing actor's recurrent work. They serve to remind the analyst, and the reader, to look across the full breadth of participants' activities as they go about growing infrastructure.

At each scale of infrastructure we have uncovered a persistent set of *concerns* for long-term sustainability. Below we describe our general formulation of common concerns, each manifest themselves in particular ways at each scale which we describe in the empirical sections of the paper as the *tensions*.

1. *Motivating contribution*: How to ensure that participants contribute in ways that are meaningful to the achievement of community infrastructure? How to ensure the continued commitment of participants over time?
2. *Aligning end-goals*: Infrastructure endeavors sustain multiple ongoing goals which often compete. Furthermore, participants in development come from different scientific traditions, with diverging purposes for participating. How are varying interests to be coordinated? Can multiple goals be satisfied while still developing effective infrastructure?
3. *Designing for use*: How to develop resources that will be adopted by users and serve in the work of research? This concern is rooted in an acknowledgement that an infrastructure without users is not infrastructure at all.

At the intersection of concerns and scales are the nine tensions in the development of long-term infrastructure (see Table 1). The concerns recur at the three scales but they carry differing implications for each. For example, the diverging end goals of individual scientists (whose focus is often conducting novel research) has implications for: i- the development of IT resources that will serve a broader community (enacting technology); ii - the division of labor between research, development and maintenance (organizing work); and iii- efforts to evaluate the activity of participants vis a vis academic reward structures (institutionalizing).

Our research goal is to understand the constitution of a general problem space for long-term infrastructure development and maintenance using the context of CI development. The identification of tensions is not to be understood as the prerogative of the analyst, but as the outcome of participants' orientations as expressed in their work. In short, not everything that could conceivably be characterized as a tension is treated as such by participants. Conversely, many of the tensions we identify are not unique to concerns for long-term sustainability. For example, in her studies of CI, Lawrence [2] has identified a *general* tension between research and development and Lee *et al.* [3] have described permeable boundaries between participants, users and community. In this paper we focus particularly on how participants articulate these tensions relative to the goal of achieving a long-term infrastructure.

3. CASES AND STUDY

We focus on four cases of data centered scientific organizations. Our cases have been chosen to highlight the different scope of the long-term within scientific information infrastructure endeavors. All projects are funded through the National Science Foundation (NSF), the federal funding body for science in the US. Roughly speaking, our cases are drawn from the earth and environmental sciences: GEON is the geosciences network, with participants drawn from over 10 disciplinary backgrounds such as geophysics and paleobotany; LTER is Long-Term Ecological Research, a consortium of 26 sites across the US that seek to collect and preserve data on time-scales that match environmental change; LEAD is Linked Environments for Atmospheric Discovery, an atmospheric science research project primarily focusing on the mesoscale weather phenonena (i.e. tornadoes) and with the hope of providing tools for real-time data analysis; WATERS draws together environmental engineering and hydrology in the development of a large-scale remote sensing network. Table 2 summarizes the cases by their targeted communities and IT development goals, by timeline, institutional affiliation and funding mechanism. These five elements are our primary material for understanding how actors approach problems of the long-term.

Our data have been collected through ethnographic study, including document collection, and have been supplemented by targeted interviews. Our primary sites for field research have been the meetings of principal investigators (PIs), designers and implementers: these are excellent sites in which key actors regularly discuss the evaluation, strategizing and planning of their projects. The time we have spent in the field with each project varies from one year (LEAD) to three years (GEON). Additionally each project has granted us access to portions of their email listservs providing a continuous stream of data; because these projects are geographically distributed email discussions often include daily decision making. Finally, in each project we have also been *participants,* contributing to aspects of planning, proposal writing, social dimensions feedback, user studies or requirements elicitation [26].

Our research was driven by grounded theory methodology [27, 28]: *iterations* of data collection combined with *testing* against substantively generated theory and comparison across our cases and with historical and contemporary studies of infrastructure [29]. Rather than the formal comparisons used to identify causal variables (as in the methods of difference and similarity c.f. J.S.Mill) we approach our cases through *constant comparisons* [25]: insight is generated by contrasting grounded

Table 2: Summary chart of four cases.

Cases	GEON (Geosciences Network)	LEAD (Linked Environments for Atmospheric Discovery)	WATERS (Water and Environmental Research Systems)	LTER (Long-Term Ecological Research)
Targeted Communities	Solid Earth Sciences	Atmospheric Sciences and Meteorology	Environmental Engineering and Hydrological Sciences	Ecological Sciences
IT Goals	Systems, Knowledge Mediation, Visualization	Workflows, Data Integration, Visualization	Instrumentation, data archiving, knowledge mediation, integration	Instrumentation, data archiving and integration
Timeline	5 yrs	5yrs	~25yrs with multiple points of funded planning, review and implementation	150yr goals, 10yr program review, 6yr site review
NSF Directorates(Division)	CISE/GEO(EAR)	CISE/GEO(ATM)	ENG/GEO(EAR)	BIO(DEB)
Funding Mechanism	ITR	ITR	MREFC	Program

categories (or codes) with multiple instances drawn from the data and historical cases. The process is iterative, leading to continuous revision of those categories. The 'scales of infrastructure' and the three 'concerns' we identify in this paper are our own analytic categories generated through the process of data collection.

We have found that many of the tensions are framed by participants in terminologies that will be familiar to social science readers; in fact, we argue that the tensions are substantially drawn from and informed by social science research. For example the terms 'career trajectory' and 'reward structure,' once the jargon of micro economics and the sociology of work (c.f. [30, 31]), have become part of the vernacular language of CI. This is because of the close relationship of social science to CI endeavors. Of the projects we study all have had the direct participation of a social scientist (ourselves and others) e.g. David Ribes worked closely with GEON for three years; LEAD and CLEANER benefited from user surveys and requirements elicitation by information studies specialists; and LTER has many members from urban sociology. Consequently it is not surprising to find CI scattered with social science concepts and terms. It is a finding of our studies that the tensions and problems of developing long-term infrastructure have at least partially been framed by actors locally enacting social science concepts and research. While we will not further develop this line of reasoning in this paper, future research will continue to track the lineages of social science findings and concepts as they innovate within CI (see Ribes and Baker [26] for a more elaborated formulation).

4. INSTITUTIONALIZATION

4.1 Project vs. facility

The majority of CI endeavors today are organized as projects. That is, they were funded under *research* awards with definite end-dates. While participants seek to develop persistent resources, in many cases there is no explicit stipulation for how these projects will be continued following the initial award.

The 'project status' and its consequences are particularly apparent for those funded under research grants. For instance, both GEON and LEAD have five year awards under the NSF's Information Technology Research (ITR) program. While these projects seek to provide stable resources to the earth and atmospheric sciences, respectively, no systematic mechanisms are available to participants for renewing the grant: this places a finite horizon on long-term activities and shifts efforts from developing stable informational resources to securing post-year five support.

In projects with uncertain futures participants have no clear sense of a preconfigured directionality for maturation [32]: what are the possible trajectories for shifting from short-term projects to stable systems and then to infrastructures? Instead, there is a continuous effort to hedge 'side bets' [33] by chasing grant opportunities and forming loose couplings [34] with institutions of science in the hope that these may later become a source of sustainable funding.

In contrast with a project status, LTER has established itself as a relatively stable 'facility'. LTER is an NSF program evaluated as a whole every ten years with clear venues for funding renewal [35]. As Karasti and Baker have noted in their analysis of LTER "the longer than usual funding cycles provide a harbor for activities such as multidisciplinary studies, network participation, and community change discussion." [36]. Achieving a longer and renewable funding cycle can mean a shift in efforts from securing funding to demonstrating effectiveness and a capacity to provide stable resources.

Returning to GEON and LEAD, a great deal of effort in such projects is invested in "politicking" as participants seek to secure something akin to facility status. As the GEON project's 5-year cycle approached closure one participant remarked that funding for science-centered infrastructure design is notoriously unstable. Specifically, rather than following a model for a persistent facility the "funding model is like any other NSF research project, i.e., relatively unstable!" (email correspondence, May 13th, 2007). Thus, throughout their five years of funding GEON participants have formed ties to the United States Geological Survey (USGS); similarly, LEAD participants have cultivated a relationship with the long-term atmospheric-data organization Unidata. USGS is a federally supported program for earth science; while Unidata has served the atmospheric research and meteorological communities for over 20 years by brokering access to data. For both projects these institutions are seen as a means to secure long-term sustainability, particularly in terms of funding but also through the 'authorization' an institution of science can offer to a nascent (and unpopulated) infrastructure.

We call this tension *projects vs. facilities*. While projects such as GEON or LEAD seek to become infrastructure, in practice, participants must operate within the finite funding window of an award (often narrow by the standards of technology enactment). The tension is manifested in the need to produce both short-term (research) products but also demonstrate long-term viability. Participants find themselves distributing their time between short-term products that can be cast as 'deliverables'[3] for upcoming project evaluation and at the same time the sustained development of stable, extensible, interoperable CI. We will see that these conditions also manifest as difficulties in effectively organizing participants' work which also shape the emerging technologies themselves.

4.2 Individual vs. community interests

In CI development, a number of 'technical' tasks must be accomplished to ensure downstream interoperability, accessibility and usability, but these tasks are not always consistent with disciplinary structures that reward contributions to domain knowledge. Most principal investigators are practicing scientists, in either a domain (e.g., geology, hydrology, meteorology) or in computer science. As individuals within academic career trajectories [30, 31] concerns emerge for meeting disciplinary criteria for advancement. In science these are usually research findings, contributions to a base of knowledge, and publications. Career trajectories are said to be institutional in that they are distributed across, for example, scientific norms ('what is or is not a contribution to hydrology?'), venues for publication ('can metadata be published?') and position within an organization (e.g. assistant professor).

CI projects are often planned to address particular 'science questions,' in the hope of ensuring contributions to the domain field [37]. For example, participants in WATERS have identified 'grand challenge' science questions driving contemporary hydrology and environmental engineering such as "How do we detect and predict waterborne hazards in real time?" Meanwhile, the infrastructure itself is intended to serve beyond any idiosyncratic agenda to the broader communities' diverse research requirements: data should be accessibly catalogued rather than hoarded; tools should support an array of general scientific research tasks rather than the particular needs of a given researcher; computing cycles should be allocated through transparent mechanisms rather than an invisible college. Since tools developed to engage particular research questions do not automatically translate to those which will support a community's long-term scientific activity [38] a tension emerges between *career rewards and community interests*: while individual needs must be met today, participants must also develop tools that in the future will serve research tasks of a community.

The tension is most often expressed as a concern for junior researchers. For example, many of GEON's participants are graduate students in the earth sciences; as part of their training geologists must generate and publish new knowledge about the earth. However, as participants in an infrastructure development project many of these students have dedicated substantial portions of their time to coding metadata for geological databases or fine tuning knowledge representations that will later assist in searches or registering data. While these are clear contributions to the informational resources of the geosciences, many earth scientists would not consider it a contribution to geoscience or, in GEON's common parlance, 'something new about the Rockies.' In the future these graduate students may have difficulty justifying their expertise as 'geological' to their doctoral committee and may appear less qualified to a hiring agency. *How to reward activities that will lead to the development of community resources today while still furthering the interests of individual scientists?*

4.3 General communities vs. specific constituencies

In any given infrastructure endeavor we should not take the communities of infrastructure as a given. As they seek to become facilities the prospecting activities of project participants lead to shifting alliances. We shall use the term *community* to refer to the general body of the domain and *constituency* to refer to the more particular groups and organizations tied to the project. For example, while WATERS is intended to serve the community of hydrologists, more practically it is organizationally tied to a consortium of hydrologists at 121 universities called CUAHSI[4]; that is their constituency. Communities are amorphous and abstract entities that come to be represented by particular research groups and scientific organizations.

Funding structures and organizational alliances shift over time as infrastructures institutionalize, and with these shifts constituencies may be completely overturned. For example, the WATERS network began in 2005 as CLEANER: Collaborative Large-scale Engineering Analysis Network for Environmental Research. CLEANER's IT resources were to be directed primarily at the environmental engineering community. However, goals for data integration and cost-sharing eventually led to an alliance between environmental engineers and hydrological scientists. There is a blurry boundary between these two groups of researchers: within NSF they are funded through distinct units, engineers tend to identify as practically oriented while scientists tend to emphasize knowledge production; however, often members of both groups cross identify, collaborate and publish in the same journals. In 2006 CUAHSI was added to the CLEANER team and the project was renamed WATERS to reflect its plans to serve both constituencies. Following this union, Katherine Lawrence's comparative survey studies showed how many hydrologists and environmental engineers were willing to cross-identify as part of the same community, however, particular constituencies were interested in access to differing databases and even the software technologies themselves would have to tailored across these: "tools will need to allow flexibility within the search function so that researchers can find data based on specific measures or variables of interest, whether they are geographic, time based, or some other metric," [39]. While WATERS was always intended to serve researchers focusing on fresh water (community) its changing configuration of partners (constituency) required adjustments in the developmental trajectories of its IT technology.

As participants shape their ties to the institutions of science their mandated constituencies also come to shift, but this also means a shift in the direction of technological development, in the adoption of community specific data standards or even vocabularies. For participants in design a significant

[3] Very often in the form of demonstration tools (or 'demos') which point to promising future capacities, but which are not of production quality (see below).

[4] Consortium of Universities for Advancement of Hydrologic Science.

transformation of the constituency means a new set of intended users and work patterns, new kinds of research questions, and new sets of data to integrate. Ties to particular constituencies also shifts research and technological development trajectories.

5. ORGANIZING WORK

5.1 Planned vs. emergent

Larger CI endeavors such as WATERS and LTER rely heavily on project management tools to define, orchestrate, and track milestones (e.g., Gantt charts, organizational charts, work breakdown structures, earned value calculations). Smaller endeavors such as LEAD and GEON have less formalized structures but still outline their trajectories and timelines, regularly referring back to their funding proposals [37]. However, the novelty of CI often defies rationalized planning; due both to the unpredictability of the underlying technology (e.g., will something that works for hundreds of users scale to thousands?) and the uncertainty about constituencies and user requirements. As a result, enormous energy can be expended on low-level technical work that never makes it into production infrastructure, or that is rejected on delivery.

This tension links back to the institutionalization of the project: future funding and targeted communities can shift with institutional couplings. For example, as LEAD has increasingly tied itself to Unidata (see projects vs. facilities, above) it has had to align its goals with that institution's mandates. While LEAD's informational resources could be useful to biologists or ecologists (as was suggested in their funding proposal), Unidata serves the atmospheric sciences. Referring to Unidata as their future 'host' a LEAD participant wrote in an email:

> *any broadening of who the infrastructure can target needs to be done within the scope of the community Unidata can and is willing to support. So that necessarily limits the "science-customer" to atmospheric and possibly geosciences. I see the host and customer as tied together - if we venture from one, we necessarily have to revisit the other (email, 2/2007)*

Plans are a form of work. They are a carefully crafted investment in time. Demonstrating a considered plan is part of a successful funding proposal and later evaluations will judge performance relative to stated plans. Conversely, infrastructure projects find themselves having to maintain flexibility in the face of unstable funding situations, emerging technologies or the requirements of their constituencies.

This tension can manifest itself as frustration with the changing priorities of the project. For example, in CI projects we can see shifting relations between IT and domain participants: as the funding from ITR (which emphasizes novel computer science research) comes to an end, projects such as GEON and LEAD have found themselves with stronger institutional ties to the domain rather than computer science. This will mean that any continuation of those projects will necessarily require a significant 'scaling down' of the IT research elements in favor of domain science. Obviously this is not only a shift in identity but also a transformation of the kinds of work done in the project, i.e. will the computer science principal investigators have any interest in renewing with a diminished IT research budget? As the future trajectory of development shifts carefully laid plans, divisions of labor or even technological trajectories may be abandoned. Members in these projects 'signed on' to participate in particular ways but emerging requirements may leave their contributions marginalized or demanding a significant reworking.

5.2 Research vs. development

Many participants in CI are scientists. As they describe themselves, scientists' personal interests, institutionalized career trajectories and community norms encourage contributions in the form of new knowledge. This applies equally to computer and domain scientists. Furthermore projects funded as research (such as those under ITR: LEAD and GEON) are partially evaluated relative to their 'science contributions.' For example, GEON's first funding proposal was rejected for leaning too far in the direction of technology deployment and with too little emphasis on geoscience research; GEON participants must be conducting geoscience research. However, committing to infrastructure design and implementation means a significant amount of effort devoted to basic technical work or implementation e.g. writing metadata, 'debugging,' and usability testing. While critical, such work is difficult to frame as science research (even for IT participants). A novel visualization tool is interesting for the hydrologist but not a contribution to the field; while it may be an effective application of IT it is not a new computational solution for computer science.

These diverging requirements within CI projects make the organization of work difficult and lead to a tension between conducting *novel research and doing the basic technical development work*. In particular, for computer science participants there is a continuous negotiation of the line between supporting tool development and slipping into a service capacity:

> *weighted too heavily toward the domain scientists, the focus overemphasizes procurement of existing technologies, and computer scientists become viewed as "merely" consultants and implementers. If the weight shifts too heavily toward computer science, the needs of end users may not be sufficiently addressed, or effort shifts too heavily toward creating new technologies with insufficient attention to stability and user support. [1: 4.3-4.4]*

For example the initial LEAD proposal did not include time and resources for requirements solicitation or usability testing. While these are crucial elements in designing systems that users will find accessible and intuitive, such activities are not research for either computer or atmospheric scientists. Because of this they easily fall from consideration; proposal reviewers may even look upon an allotment of resources to such activities as outside the range of research funding. Over time LEAD came to find additional resources and personnel to fill these roles, but the effort required to do so is not clearly credited by existing modes of evaluation. Dividing time between research and development often becomes a problem for the administrative or managerial agents in CI: *how to organize the work such that participants are both conducting research and performing the more basic technical tasks required to deliver production CI systems?*

5.3 Development vs. maintenance

Utilities cannot be available one day and not another: telephones do not crash; power supplies do not fluctuate; and clocks do not halt (in general). This stability is not only a function of technology, but of its everyday human work of repair and maintenance [40]. Similarly, a computational tool supporting everyday scientific work must be reliable across time, it must be maintained. We can further subdivide 'technical work' in CI projects into development and maintenance: building CI itself is development but, once built, the operation of these systems must be sustained. We have already noted that development work falls behind research activities: participants describe the development of new knowledge as their primary interest and form of recognition. Maintenance work is usually even less considered; in projects without long-term support (such as GEON and LEAD) maintenance is well outside the scope of initial planning. This said, even within a five-year cycle, maintenance, repair and upgrading will be required. Computing systems today are continuously tweaked, updated or modified. To keep up, CI requires maintenance. A tension emerges between the need to develop new infrastructural resources and the continuous work of updating the existing resources so as not to fall into disrepair.

For example, LTER has the mandate to support today's ecological research but also to think ahead for its future (thus 'long-term ecological research'). The information managers in LTER have described a struggle to balance the 'to do lists' of LTER scientists as they go about accessing data and conducting research, and thinking about the development of new database technologies to preserve the archive. The tension is often framed as a matter of time management. An LTER information manager, tasked with facilitating access to data but also with planning future archival systems, expressed this as 'a balance':

> *on the one hand I know we have to keep it all running, but on the other LTER is about long-term data archiving. If we want to do that we have to have the time to test and enact new approaches. But if we're working on the to-do lists we aren't' working on the tomorrow-list. (Workgroup discussion 10/05).*

The tension described here is of time management, but also of the differing valuations for these kinds of work: the implicit hierarchy places scientific research first, followed by deployment of new analytic tools and resources, and trailed by maintenance work. Trigg and Bødker have characterized a dialectic relationship between development and maintenance by identifying organizational members they call 'tailors' who in the activity of maintaining a technical system also adjust the organization of work around it [41]. To understand *maintenance* Trig and Bødker tie the work of organizational change to the technological; they argue that it is through minor adjustments to both that technological development is conducted. The concept of tailoring partially reflects the activities of LTER information managers but it fails to capture the work of negotiating competing requirements within information infrastructure design. While in an ideal situation development could be tied to everyday maintenance work, in practice maintenance work is often invisible and undervalued. As Leigh Star notes, infrastructure becomes visible upon breakdown and only then is attention directed at its everyday workings [27]. *How to support the distribution of work across research, development and maintenance?*

6. ENACTING TECHNOLOGY

6.1 Inclusion vs. readiness

CI is intended to serve a community. For example WATERS serves both hydrologists and environmental engineers and GEON serves the geosciences, i.e. paleobotanists, metamorphic petrologists and geophysicists, to name only a few. The goal is to develop an 'umbrella' which will serve the research needs of scientists but also help foster communication and collaboration. In practice, however, participants find that various fields differ in their 'readiness' for CI. For example, seismologists and geophysicists have long traditions of using remote instrumentation and shared databases while paleobotanists and metamorphic petrologists are 'field scientists,' more familiar with smaller-scale data collection conducted within their own research teams.

Under such constraints a tension emerges between goals for *inclusion and readiness* across differently prepared scientific fields. This is true both from the 'hard' and 'soft' foundations of CI [42]: computing and network capacity on the one hand and on the other the community's experience collaborating or willingness to share data. From the perspective of CI deployment, the data of some fields are more readily available than others for federation. GEON's metamorphic petrologists describe their data as distributed in a multitude of publications; they must 'begin' with a process of digitization, database design and metadata creation. Meanwhile geochemists are much further along, with parallel integration effort such as 'EarthChem' which was initiated in 2003 with the goal of interoperating three major databases from that field[5]. Such efforts place the geochemical community in a better position to take advantage of the high-end computing resources of infrastructure projects such as GEON. However if development efforts are directed at the 'ready' community of geoscientists an uneven development of CI may result, leaving metamorphic petrologists well outside the 'inclusive umbrella infrastructure.' Inclusion is also a matter of attempting to evenly distribute CI development.

A short-term consequence of uneven development of CI is the marginalization of participants (such as metamorphic petrologists). Star and Ruhleder (1994) have described such actors as the 'orphans' of infrastructure. A long-term consequence could mean an increased 'digital divide' as development resources are funneled to those with already established technical bases; Edwards *et al.* (2007) have warned of the danger that infrastructure may be 'captured' as resources accumulate among disciplines with long histories of technical investment.

6.2 Today's requirements vs. tomorrow's uses

Designing an information infrastructure is a visionary process. While CI tools are intended to 'support the work of scientists' it is not only for today's users, but also tomorrow's. Information technology is changing rapidly as are the scientific practices and instruments that tie to those technologies. Designing to meet 'the needs of users,' then, is a moving target, notoriously difficult to represent through the self-characterizations of users, through surveys, or ethnographies [44].

For example, LEAD participants collected user requirements through surveys and interviews. The results from atmospheric scientists showed that members of that community were most interested in tools that would support their use of, for example, spreadsheets to manipulate data. Today spreadsheets remain a staple for atmospheric scientists in arranging and manipulating their data. However, for a high-end technology development project such as LEAD this finding is directly at odds with the

[5] PetDB, NAVDAT,GEOROC [43].

vision of revolutionizing scientific practice or conducting cutting edge computer science. GEON participants made similar assessments: they noted that requirements solicitation is not usually a forward looking activity, when it is, expectations are unrealistic:

> *There are two results when you ask a user what they want: they will either tell you something completely banal that's completely uninteresting to develop or they will ask something in the range of science fiction, well beyond the current state of the art. (interview 11/2003)*

Put simply, the majority of domain scientists inhabit computing environments that operate on the average desktop (word processors, browsers, spreadsheets) along with a handful of specialized domain applications. When asked about a future trajectory of technological development they respond first with their everyday computing pains, and if pushed will generate digital utopias. They do not reside at the forefront of computer science and are unable to articulate realistic contemporary applications.

However, you cannot altogether abandon requirements elicitation. Designers are increasingly well-versed in the danger of designing systems independent of involvement with users. Systems designed without consideration for user's work patterns (paraphrasing Suchman 'tossed over the design wall'[45]) often leave large gaps between work routines and technical capacities. A tension emerges between the constituencies' demands and forward looking development. Respecting current work practices is a key feature of successful system design, but this leaves designers with unclear trajectories for 'transforming scientific practice.' Moreover, this tension dovetails with that of projects vs. facilities: projects must demonstrate short term results to appear viable, but developing truly useful and stable applications is a long-term endeavor.

6.3 Research vs. production quality systems

The distinction between research and production quality systems is well known within information technology circles [32, 45]**.** Research systems are the bread and butter of applied computer science: they serve as platforms for testing novel technologies but they are renowned for their instability, poor documentation and relative unfriendliness to the user. For example, experimental tools within research systems are often called 'demos' and, as one GEON programmer noted, they are notoriously limited in their functionalities:

> *they are pretty, they do new and amazing things, and you have to stand by the user the whole time to make sure he doesn't do a single thing you haven't supported. Oh yeah, and it usually turns out that's mostly what they want to do. (interview 11/2003)*

In contrast, production quality systems are stable and reliable, they have been 'hardened' through user testing, and often designed with concerns for usability by creating intuitive interfaces and thorough documentation. However, as described above, the development of such stable 'tried-and-true' applications is often not considered a computer science contribution. Under such constraints a tension emerges between producing systems that support computer science research or creating stable reliable and long-term resources for a community of everyday users.

The rhetoric of cyberinfrastructure usually emphasizes changing practices and methods for scientific research; however, this underemphasizes the instability of novel applications. As one LEAD developer noted of meteorologist users: "they're willing to give our stuff a try, but the second time it breaks while they're using it they'll never come back […] What they want is reliability, and ubiquitous reliability is the critical part" (workgroup discussion 12/06). The consequence may be that disabused and disillusioned users will be unwilling to migrate to new infrastructures, leaving a landscape dotted with promising but unpopulated computational architectures.

We have already described the institutionalized reward systems for researchers and the difficulties this raises in organizing development or maintenance work. The resulting dynamic also affects the emerging technology itself as systems are geared to serve further IT research *or* practical uses for domain scientists.

7. CONCLUSION

There is something paradoxical in a long-term plan for information technology. We think of IT as changing at a rapid and ever increasing pace; yesterday's novel solutions quickly become today's staple resources and even faster seem to become tomorrow's relics. This is the challenge of the *long-term today*, of an effective transition from one-off applications and prototypes to stable and usable informational facilities. We still have only a handful of examples that persist beyond a decade.

Returning to the "Millennium Clock," we are now in a position to understand Stuart Brand's concept of "the long now." The long now is a conceptualization of time which demands that sustainability become *today's* consideration. The designers of the clock plan for it to endure ten thousand years, but, between then and now, the clock must be built, wound, maintained and housed. The time between today and ten thousand years requires the sort of thinking Brand calls 'the long now.' In thinking in the long now the clock's designers have regularly worked across the scales of infrastructure: building the clock itself, with its wear-resistant gears and hardened case, is a technological question; winding the clock, be it annually or by decade, is a matter of human organization; and housing the clock over centuries can only be an institutional responsibility. But before all this, planning a long-term is the work of today's design: "*Three years we've been working on building* a ten-thousand-year clock" (ibid.). These are the tensions of the long-term today which we have sought to capture and characterize in this paper.

The long-term aspects of infrastructure will require a systemic response, or as Lee *et al.* (2006) describe it, the creation of "human infrastructure" to operate cyberinfrastructure. This concept draws attention to often invisible participants and their work of enactment, maintenance and upkeep. However, while the work of design and development is 'human,' the challenges are more comprehensively described as technical, organizational and institutional. *In considering design and enactment of infrastructure it is best to address 'hard and soft' foundations hand-in-hand, they are usually more intimately entwined than any raw distinction would suggest.* This insight is reflected in the terminology which has come to populate the social studies of infrastructure, concepts such as 'sociotechnical' [6] which link a broader context to local technology development, or 'technoscience' [46] which discourage distinctions between scientific practice and the instrumentation which enable that research. The difficulties of long-term design are reflected in the nine tensions framed by participants that we have organized around three concerns (motivating contributions, aligning end goals and designing for use) and three scales of infrastructure (institutionalizing, organizing work, and enacting technology). Just as we have seen the links between, for example, instability of funding and the enactment of experimental systems, actors formulate solutions which simultaneously address multiple concerns across the scales. In each case, strategies to manage the tensions will have implications in terms of the appearance and organization of resulting information infrastructure.

We remain agnostic with respect to whether these choices are good or bad, and this paper is not meant to offer normative guidance on the "right way" to produce CI. Rather, our goal is to remind developers and users of CI that such choices do exist and to begin exploring their consequences. Our intent in this paper is to illuminate the problem-space in which choices about information infrastructure design are made and by doing so to provide a theoretical framework for conducting research on the emergence of CI.

Breakdowns in CI are likely to occur and when they do, will be of great interest, in the same way that breakdowns in other kinds of infrastructure have been notable (e.g., the Three Mile Island nuclear accident in 1979). As in the case of previous infrastructure failures, the seeds of CI failure will undoubtedly lie within lapses or gaps in the organizations designed to maintain and operate CI. To minimize or avoid these failures, then, will require deep understanding of the ways in which CI designers, developers and users strategize to resolve tensions as they go about the work of design and enactment. This said, the most pervasive concern within funding agencies, amongst designers and for future users is not of breakdown, but in the failure to receive adoption. Beautiful but empty informational corridors are not infrastructure at all.

This paper has focused on actor's work of problem identification. However, just as participants are regularly engaged in articulating problems for long-term development (often as 'tensions') they are also seeking solutions. For example, we have briefly discussed how in the face of short-term funding CI projects will attempt to transition to facilities by forming alliances with the persistent institutions of science in their domain fields. Thus, the constitution of a problem space is coupled to the enactment of its solutions; we call this a *strategy of the long-term*. Our future research will focus on how actors formulate responses in the face of the problems they have identified.

8. ACKNOWLEDGEMENTS

Recognition is given to the Comparative Interoperability Project team members Geoffrey C. Bowker, Florence Millerand and particularly Karen Baker whose insight and collaborative work is integral to this effort. We'd also like to thank Ann Zimmerman for her feedback and Katherine Lawrence for her previous work with LEAD and WATERS. This work was supported by the National Center for Supercomputing Applications (NCSA) Partner Project; NSF/SBE/SES Human Social Dynamics grant #04-33369 (Interoperability Strategies for Scientific Cyberinfrastructure: A Comparative Study); and NSF/SES grant #0525985 ('The Standardized Revolution of Science: Building Cyberinfrastructure for the Geosciences').

9. REFERENCES

1. Atkins, D.E.C., *Revolutionizing Science and Engineering Through Cyberinfrastructure: Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure*. 2003, National Science Foundation.
2. Lawrence, K.A., *Walking the Tightrope: The Balancing Acts of a Large e-Research Project.* Computer Supported Cooperative Work, 2006. **15**(4): p. 385-411.
3. Lee, C.P., P. Dourish, and G. Mark, *The human infrastructure of cyberinfrastructure.* Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work, 2006. **NY, NY**: p. 483-92.
4. Star, S.L. and K. Ruhleder. *Steps Towards and Ecology of Infrastructure: Complex Problems in Design and Access for Large-Scale Collaborative Systems*. in *Proceedings of the Conference on Computer Supported Cooperative Work*. 1994. Chapel Hill: ACM Press.
5. Brand, S. and J.i. Brockman, *The Clock of the Long Now: A Talk with Stewart Brand.* Accessesed on Jan.10.07 http://www.edge.org/3rd_culture/brand/, 1998.
6. Edwards, P., *Infrastructure and Modernity: Force, Time, and Social Organization in the History of Sociotechnical Systems*, in *Modernity and Technology*, T.J. Misa, P. Brey, and A. Feenberg, Editors. 2003, The MIT Press: Cambridge, MA.
7. Zimmerman, A., *New Knowledge from Old Data: The Role of Standards in the Sharing and Reuse of Ecological Data.* Science, Technology & Human Values, Forthcoming.
8. Rodden, T., J.A. Mariani, and G. Blair, *Supporting Cooperative Applications.* Computer Supported Cooperative Work, 1992. **1**: p. 41-67.
9. Sheth, A.P., *Changing Focus on Interoperability in Information Systems: From System, Syntax, Structure to Semantics*, in *Interoperating Geographic Information Systems*, E. Goodchild, Fegeas, Kottman, Editor. 1999, Kluwer Academic Publishers: Boston.
10. Argyris, C. and D.A. Schon, *Organizational Learning II: Theory, Method, and Practice*. 1996, Reading, Mass: Addison-Wesley Pub. Co.
11. Fountain, J.E., *Building the Virtual State: Information Technology and Institutional Change*. 2001, Washington, D.C.: Brookings Institution Press.
12. Heath, C., et al., *Unpacking Collaboration: The Interactional Organisation of Trading in a City Dealing Room.* Computer Supported Cooperative Work, 1993. **3**: p. 147-165.
13. Lougher, R. and T. Rodden, *Supporting long-term collaboration in software maintenance.* Proceedings of the conference on Organizational computing systems, 1993: p. 228-238.
14. Lindstaedt, S. and K. Schneider. *Bridging the gap between face-to-face communication and long-term collaboration*. in *Proceedings of the international ACM SIGGROUP conference on Supporting group work: the integration challenge*. 1997. New York, NY: AC Press.

15. Finholt, T.A., *Collaboratories.* Annual Review of Information Science and Technology, 2004. **36**: p. 73-107.
16. Olson, G.M. and J.S. Olson, *Distance Matters.* Human-Computer Interaction, 2000. **15**(2-3): p. 139-179.
17. Cummings, J. and S. Kiesler, *Coordination and success in multidisciplinary scientific collaborations.* Social Studies of Science, 2005. **34**(?): p. 1-20.
18. Bowker, G.C., *Memory Practices in the Sciences*. 2006, Cambridge, MA: MIT Press.
19. Meyer, J. and B. Rowan, *Institutionalized Organizations: formal structure as myth and ceremony.* American Journal of Sociology, 1977. **83**(340-63).
20. March, J.G. and J. Olsen, *Rediscovering Institutions*. 1989, New York: The Free Press.
21. Selznick, P., *Institutionalism "old" and "new".* Administrative Science Quarterly, 1997. **41**: p. 270-7.
22. David, P.A. and M. Spence, *Towards insitutional infrastructures for e-science: The scope of the challenge.* OII Research Report No.2 (Retrieved April 6, 2007) from http://www.oii.ox.ac.uk/resources/publications/RR2.pdfm, 2003.
23. Pfeffer, J. and G. Salancik, *Organizational Decision Making as a Political Process: The Case of a University Budget.* Administrative Science Quarterly, 1974. **19**(2): p. 135-151.
24. Callon, M. and B. Latour, *Unscrewing the Big Leviathan: how actors macrostructure reality and how sociologists help them to do so*, in *Advances in Social Theory and Methodology: Toward an Integration of Micro- and Macro-Sociologies*, K. Knorr-Cetina and A.V. Cicourel, Editors. 1981, Routledge: Boston, Mass.
25. Glaser, B.G., *Theoretical sensitivity: advances in the methodology of grounded theory*. 1978, Mill Valley, CA: Sociology Press.
26. Ribes, D. and K.S. Baker, *Modes of Social Science Engagement in Community Infrastructure Design*, in *Proceedings of Third International Conference on Communities and Technology*, Steinfield, et al., Editors. 2007, Springer: London. p. 107-130.
27. Star, S.L., *The Ethnography of Infrastructure.* American Behavioral Scientist, 1999. **43**: p. 377-391.
28. Clarke, A., *Situational analysis: grounded theory after the postmodern turn*. 2005, Thousand Oaks: SAGE Publications.
29. Bowker, G.C., et al., *Towards Information Infrastructure Studies: Ways of Knowing in a Networked Environment*, in *International Handbook of Internet Research*, J.D. Hunsinger, M. Allen, and L. Klastrup, Editors. forthcoming, Springer.
30. Fairweather, J.S., *Faculty Reward Structure: Toward institutional and professional homogenization.* Research in Higher Education, 1993. **34**(5).
31. Kling, R. and L.B. Spector, *Rewards for scholarly communication*, in *Digital scholarship in the tenure, promotion, and review process*, D.L. Anderson, Editor. 2003, M.E. Sharpe Inc: Armonk, NY.
32. Weedman, J., *The Structure of Incentive: Design and Client Roles in Application-Oriented Research.* Science, Technology & Human Values, 1998. **23**(3): p. 315-345.
33. Becker, H.S., *Notes on the Concept of Commitment.* The American Journal of Sociology, 1960. **66**(1): p. 32-40.
34. Weick, K.E., *Educational organizations as loosely coupled systems.* Administrative Science Quarterly, 1976. **21**: p. 1-19.
35. Hobbie, J.E., et al., *The US Long Term Ecological Research Program.* BioScience, 2003. **53**(2): p. 21-32.
36. Karasti, H. and K.S. Baker. *Infrastructuring for the Long-Term: Ecological Information Management*. in *37th Annual Hawaii International Conference for Sytem Science Proceedings*. 2004: HICSS37, IEEE Computer Society.
37. Ribes, D. and G.C. Bowker, *Organizing for Multidisciplinary Collaboration: The Case of GEON*, in *Science on the Internet*, G.M. Olson, J.S. Olson, and A. Zimmerman, Editors. forthcoming, MIT Press: Cambridge.
38. Greenberg, S., *Personalizable Groupware: Accomodating Individual Roles and Group Differences*, in *In Proceedings of the European Conference of Computer Supported Cooperative Work (ECSCW '91), pp.17-32*. 1991, Kluwer Academic Press: Amsterdam.
39. Finholt, T.A. and J. Van Briesen (co-chairs). *WATERS Network Cyberinfrastructure Plan: The WATERS Network Project Office Cyberinfrastructure Committee*. 2007 [cited 2007 Aug. 27]; Unpublished technical report, please contact Thomas A. Finholt, finholt@umich.edu, or see http://cleaner.ncsa.uiuc.edu/home/].
40. Graham, S. and N. Thrift, *Out of Order: Understanding Repair and Maintenance.* Theory, Culture & Society, 2007. **24**(3): p. 1-25.
41. Trigg, R. and S. Bødker. *From implementation to design: tailoring and the emergence of systematization in CSCW*. in *Proceedings of the 1994 ACM conference on Computer supported cooperative work*. 1994. Chapel Hill, NC: ACM Press.
42. David, P.A., *Towards a cyberinfrastructure for enhanced scientific collaboration: providing its 'soft' foundations may be the hardest part.* OII Research Report No.2 (Retrieved April 6, 2007) from http://www.oii.ox.ac.uk/resources/publications/RR2.pdfm, 2004.
43. Walker, J., et al., *EarthChem: International Collaboration for Solid Earth Geochemistry in Geoinformatics.* Eos Trans. AGU Fall Meet Suppl., 2005. **86**(52).
44. Jirotka, M. and J. Goguen, *Requirements Engineering: Social and Technical Issues*. 1994, New York: Academic Press.
45. Suchman, L., *Working Relations of Technology Production and Use.* Computer Supported Cooperative Work, 1994. **2**(30): p. 21-39.
46. Latour, B., *Science in action: how to follow scientists and engineers through society*. 1987, Cambridge, Mass.: Harvard University Press. 274.