Planning infrastructure for the long-term: Learning from cases in the natural sciences

David Ribes[1], Thomas A. Finholt[1]

[1] School of Information, University of Michigan

Email address of corresponding author: dribes@umich.edu

Abstract. E-social science involves the long-term provision of human services and information technology to support cyber-enabled forms of social science research. In implementing across an extended timescale, the social sciences will confront difficulties comparable to those observed in the natural sciences, including sustainability, extensibility, and usability of systems. In this paper we draw on our prior and ongoing comparative social studies of cyberinfrastructure (CI) in the natural sciences to inform policy, planning and implementation of e-social science. We argue that in their efforts to build CI, social scientists can learn from the experiences of natural science communities. Specifically, we identify three primary concerns and describe exemplary tensions social scientists are likely to encounter as they design CI for the long-term. We conclude by exploring the potential for a "latecomer" advantage where social scientists may overcome tensions of designing CI for the long-term through delayed adoption as socio-technical arrangements stabilize.

1. Introduction

In planning for e-social science we are speaking of infrastructure: the provision of human services and information technology to support cyber-enabled forms of research in the social sciences. The notion of e-social science evokes images beyond 'a demonstration,' a 'prototype' or an isolated 'application,' rather, e-social science is intended to be a persistent, ubiquitous and reliable environment. In designing infrastructure for e-social science, participants will be asked to think in the long-term. However, in implementing across an extended timescale we expect the social sciences will confront difficulties in building sustainable infrastructure which are similar to those observed in the natural sciences, including: achieving stable funding; support for evolving systems; and sustaining engagement by participating scientists.

In this paper we draw on our prior and ongoing comparative social studies of cyberinfrastructure (CI) for the natural sciences to inform policy, planning and implementation of e-social science. In the first section of this paper, we explore three primary concerns and describe exemplary tensions encountered by participants in CI as they go about designing for the long-term. In the second section of this paper we argue that in building infrastructure social scientists should draw insights from the experience of natural scientists building cyberinfrastructure. Through comparative studies of multiple natural science CI projects we have catalogued the difficulties that participants in these projects describe as they go about developing long term infrastructure to support scientific work. We conclude by hypothesizing that awareness of CI development in natural science communities may provide a 'latecomer advantage' for social scientists. Specifically, through delayed adoption of CI, we expect that social scientists may benefit to the extent that technical and socio-technical arrangements have matured and stabilized – and therefore reduce some of the tensions associated with the long-term design and use of CI.

Our research has ethnographic goals (Star 1999): we seek to convey the tensions in cyberinfrastructure design by capturing the orientation and actions of participants. Our data are drawn from observing the *in situ* work of designers, project managers, scientists and other participants in CI projects. We approach our cases through the constant comparisons of grounded theory (Glaser 1978). The method is intended to generate insight by contrasting cases in an iterative process of research.

2. Methods and Cases

We focus on four cases of data-centered scientific organizations, summarized in Table 1. 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 ten 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 phenomena (i.e., tornadoes) and with the hope of providing tools for real-time data analysis; and WATERS draws together environmental engineering and hydrology in the development of a large-scale remote sensing network. Table 1 categorizes the cases by their targeted communities and IT development goals, by timeline, institutional support 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, designers and implementers: these are excellent sites in which key actors regularly discuss the performance, strategies, and plans for 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

Table 1: 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
Institutional support (NSF Directorate and Division)	CISE/GEO(EAR)	CISE/GEO(ATM)	ENG/GEO(EAR)	BIO(DEB)
Funding Mechanism	ITR	ITR	MREFC	Program

email listservs, providing a continuous stream of data; because these projects are 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 and user requirements elicitation (Jirotka and Goguen 1994).

Our research was driven by grounded theory methodology (Clarke 2005; Strauss 1993) iterations of data collection combined with testing against substantively generated theory and comparison across our cases and with historical and contemporary studies of infrastructure (Ribes and Baker 2007). 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 (Glaser 1978): Insight is generated by contrasting grounded 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 tensions we identify are not unique to concerns in long-term sustainability. For example, in her studies of CI Lawrence (2006) has identified a general tension between research and development and Lee *et al.* (2006) have described fuzzy boundaries between participants and community. In this paper we focus particularly on how participants articulate such tensions relative to the goal of achieving a long-term infrastructure.

3. Concerns and tensions in CI development

Infrastructure stretches across multiple scales of action: institutionalizing; organizing work; and enacting technology. At the institutionalizing scale, builders of infrastructure seek to provide service to communities at national and international scales. At the scale of organizing work, creation of infrastructure is a matter of care and maintenance, or as sociologist Leigh Star reminds us, one person's infrastructure is another person's daily routine of upkeep (1994). Finally, at the scale of enacting technology, developers of infrastructure want to design and deploy durable resources to support work, automate tedious tasks and enable collaboration. 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' (Edwards 2003).

At each scale of infrastructure we have uncovered a persistent set of concerns for long-term CI development. Below we describe our general formulation of these concerns and later we describe how these manifest themselves in practice through exploration of three exemplary tensions[3]:

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] For a fuller treatment please see (Ribes and Finholt 2007) in which identify and describe nine tensions at the intersections of the scales and the concerns, as participants seek to design infrastructure for the long-term. In this paper our concern is how we can articulate difficulties in developing information infrastructure found in the natural sciences to enable sustainable arrangements in e-social science.

3. Designing for use: How to develop resources that will achieve uptake 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.

In each of the following sub-sections we examine exemplary tensions at the intersection of scales of action and concerns. Specifically, the first tension, termed *project or facility*, arises where participants struggle to produce enduring facilities (institutionalizing) through the mechanism of problem-specific projects of limited duration (alignment of end-goals). The second tension, termed *development or maintenance*, reflects the need for certain tasks to be accomplished (organizing work) that are not necessarily consistent with reward structures (motivating contributions). And finally, the third tension, termed *inclusion or innovation*, emerges where participants seek to develop systems that meet the needs of future users (enacting technology) which conflicts with goals to conduct novel computer science research; these systems are notoriously unstable, and many scientific communities may not be ready to adopt them (designing for use).

3.1 Project or facility

Cyberinfrastructure has institutional goals (David and Spence 2003): it is intended to serve public purposes through the support of research; to endure beyond particular research projects or teams; and, is linked to governance and sustained state funding. Working to achieve such goals is not only a matter for policy it also a practical concern for participants.

The majority of CI endeavors in the US are organized as projects (Weedman 1998). That is, they are they are funded under research awards with definite ending dates. While participants seek to develop persistent resources, in many cases there is no explicit stipulation for how these projects will be continued following an initial award. Under such constraints a tension emerges between short-term products that can be cast as 'deliverables' (for upcoming project evaluation and reviews) and sustained development of stable, extensible, interoperable CI.

For example, both LEAD and GEON were funded under the ITR[4] program. These ambitious projects had goals of conducting basic research (in computer science, but also atmospheric and geo -science, respectively), but also of deploying community-wide infrastructure. The five year funding window in these projects enables collaborative relationships to grow, and for cutting-edge computer science experiments to mature, however participants regularly complain that the "you don't build infrastructure in five years," (all-hands meeting, 11/2002). When pressed, principal investigators will describe these endeavors as 'prototypes,' and note that these research projects did not receive funding for usability testing, extensive software hardening or even community outreach.

While such projects seek to become infrastructure, in practice, participants must operate within the finite funding window of an award. A great deal of effort in such projects is invested in attempting to secure long term funding and to align technology development with both short-term deliverables and long-term infrastructure.

---

[4] Information Technology Research. An NSF cross-directorate program which ran between 2000-5. Large ITRs awarded grants that ran 5 years, with budgets between $US10-15M.

3.2 Development or maintenance

Infrastructure design and development is a matter of practical work (Bowker, Timmermans and Star 1997). While e-social science and CI promise great advances, much of the work of implementing these systems is often mundane and routine. In the organization of work during the development process a tension emerges between conducting novel research and doing the basic technical work necessary for producing stable, interoperable systems.

We can subdivide 'technical work' in CI projects into development and maintenance[6]: building the CI itself is development work but, once built, the operation of these systems must be sustained (Trigg and Bodker 1994). Computing systems today are continuously tweaked, updated or modified. To keep up, CI requires 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). Similarly, a computational tool supporting everyday work must be reliable across time, it must be maintained. Jewett and Kling have described maintenance as the "hidden soft underbelly of computerization," (Jewett and Kling 1991:270).

A tension emerges between the need to develop new infrastructural resources and the continuous work of updating the existing systems so as not to fall into disrepair. Who will do the detailed work of development and deployment, and who will do the mundane work of maintenance and repair? Scientists' personal interests, institutionalized career trajectories and community norms encourage contributions in the form of new knowledge; they are, after all, researchers. However, committing to infrastructure design and implementation means a certain amount of basic technical work, e.g. writing metadata, 'debugging' and usability testing. While critical, such work is difficult to frame as computer science research (for either a domain or computer scientist).

Furthermore, for computer scientists and information technologists there is a thin but important line between providing the support necessary for the research and slipping into a service capacity. This is particularly a problem for the administrative or managerial agents in CI design projects: how to organize the work in CI projects such that participants are both advancing in their careers (conducting research) and performing the more basic technical tasks required to deliver production CI systems?

3.3 Inclusion or innovation

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

---

[6] Thus, this tension actually 'pulls' in three directions: scientists wishing to conduct research (in the domain or computer), information technologists implementing novel systems, and the maintenance work which is often left aside or becomes the work of a technical staff and information managers.

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 to adopt novel IT* across differently prepared scientific fields. This is true both from the ‘hard’ and ‘soft’ foundations of CI (David 2004): 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[7]. 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.

4. The merits of borrowing: Drawing insights from the natural sciences for e-social science

In this section we draw on concepts from evolutionary economics (largely the work of Thorstein Veblen) to frame how social science can learn from the experience of CI development within natural science communities. In particular, we draw upon the concept of “latecomer advantage” to suggest that social scientists may benefit from delayed adoption of CI, such as through the emergence more effective socio-technical arrangements. We believe social scientists can use their latecomer advantage to mitigate the *project or facility*, *development or maintenance*, and *inclusion or innovation* tensions described in the previous section

In his comparative study of nineteenth century industrialization in Great Britain and Imperial Germany, Veblen (1915) demonstrated how Germany’s initial ‘backwards’ social and technological development later came to be an advantage. By following Britain’s lead and learning from their mistakes, Veblen argued that Germany gained a long-term advantage through the adoption of key technologies at more mature stages of development. Veblen’s canonical example concerns railway development. The British, first to extensively innovate rail transportation for cargo, adopted a gauge width for their tracks which came to be considered narrow, thus unnecessarily reducing transport speed and demanding eccentric design solutions to stabilize cars. Germany, without the ‘inertia’ of a heavy investment in the physical capital of narrow tracks was able to adopt a wider, more advantageous, gauge. While Britain’s investment in early industrial technologies and social organization eventually left them mired in obsolete technology, Germany was able to study the progress of its more advanced neighbor and selectively appropriate capital and managerial techniques. As Veblen phrases it, England took a ‘penalty for taking the lead’ during industrialization and later faltered relative to other nations; Germany was able to draw on ‘the merits of borrowing’ by beginning their industrialization from a more sophisticated stage of technical development and with a greater knowledge of human organization.

---

[7] These databases are PetDB, NAVDAT,GEOROC, see (Walker, Lehnert, Hofmann, Sarbas and Carlson 2005) for more details on this integration project.

In contemporary studies this phenomena has come to be known as the latecomer advantage. For example, Japan's innovation of lean production practices (or flexible assembly) in automobile construction took the latecomer advantage and 'leapfrogged' over U.S. assembly line mass production. A linear model of technical advance would have predicted that the US (a clear world leader in the manufacture of cars during the 1950's) would have innovated flexible assembly. However as Womack *et al.* argue (1990), lean production techniques were a specific historical response to the small size of the post-war automobile and truck market in Japan, leading to the need to produce multiple products on the same line with great attention to efficiency. Thus, 'backwards' Japan is said to have been largely able to draw on the insights, technology and organization of mass production, but then leapfrog directly into flexible assembly through a locally tempered appropriation, thereafter leaving competitors scrambling to adopt their novel methods[8].

In Veblen's cases 'the advantages of backwardness' are largely *unintended consequences*. There was no plan, or intent, on the part of German engineers to wait for a stable technical and organizational order so as to permit an advantageous leap onto the bandwagon of industrialization. The advantage of backwardness was contingent. Similarly this is the case with e-social science. Historical circumstances -- including levels of funding, the penetration of informational technologies and automated instrumentation – have largely distinguished the trajectories of natural and social sciences[9]. Consequently, within the US, projects for the development of CI in the natural sciences have received greater attention at the levels of policy, funding and technical development. However, as we begin to mobilize e-social science ventures this provides an opportunity to inspect previous investments, leveraging both technical developments but also gleaning organizational, administrative and managerial insights; e-social science should actively seek to draw on the merits of borrowing and latecomer advantages.

We do not wish to suggest a trajectory of 'rationalization' for CI. It is not our argument (as has been read from Veblen (c.f. Gerschenkron 1962)) that latecomers are able to apply technologies without the 'impurities' of inherited social organization and thus 'more rationally.' In fact a more nuanced reading of Veblen indicates that this was not his position:

> The borrowed elements have invariably been assimilated, drawn into the cultural system and so combined and shaped to its purpose as to have led to an unbroken evolution of a scheme peculiar to these (hybrid) peoples and their needs, rather than to the substitution of a scheme from outside or a piecing-out of the scheme of things into which it is intruded (Veblen 1915: 20)[10]

---

[8] Max Weber identified a similar technical outcome, presumably independent of Veblen's work: "This is the same general phenomenon as when areas which have highly developed gas illumination works or steam railroads, with large fixed capital, offer stronger obstacles to electrification than completely new areas which are opened up for electrification," (1978:III, 18, 987) as quoted in (Schneider 1971).

[9] It beyond the scope of this paper to address how these funding asymmetries came to be; King notes that within the US NSF purview of social science came relatively late with the creation of an office in 1957 and a directorate in 1961 (King 1998). England tracks some of the controversies between congress, natural scientists and the national science board as NSF came to fund social science (England 1982; Kleinman and Solovey 1995). However this particular gap is not as seminal as it may initially appear. Within any given 'spectrum' of sciences, developmental gaps emerge between 'information rich and poor' subfields (Nentwich 2003). For example, from the perspective of physics and more recently bioinformatics, the environmental sciences have less experience with technology mediated collaboration, standardized data collection, remote instrumentation or data sharing. At a finer granularity, within the earth-sciences, we have already discussed the variable 'readiness' of GEON's participating disciplines for adopting high end CI.

[10] Within organizational studies, rationalization has always had an ambivalent relationship to effectiveness, demonstrated as early as Weber who argued that rationalized legal systems could produce stagnant 'devices,' or venues for legal action

Thus Germany did not 'rationalize' rail, rather it appropriated the technology in an informed manner congruent with its own technical inheritance and political culture. This said, a culturally specific appropriation does not forestall a softer reading of Veblen that takes Germany, for example, as able to draw insights and learn from observing the historical case of Britain. Similarly the tactic for e-social science should be framed as a local appropriation of findings from the experience of CI but according to the specific disciplinary schemas of the various social sciences.

Veblen articulated two mechanisms by which latecomers may gain advantage: i - obsolescence of machinery through improvement and ii - through a transformation in the 'technological situation.' To understand both mechanisms we must first understand what today is called the inertia of the 'installed base': "economical, technical, and organizational investments in the existing […] infrastructure." (Monteiro 1997:230). These investments preclude, or at least discourage, radical changes in base technologies which, in turn, shape developmental trajectories of emerging technologies. In the case of gauges British engineers had themselves realized the shortcomings of the narrower rail; however, the capital and industrial investment in tracks, locomotives and switching facilities forestalled any economical national transition to the more advantageous technology: "while this item of 'depreciation through obsolescence' has been known for some time, it has not even in the most genial speculative sense come up for consideration as a remediable defect" (Veblen 1915). Meanwhile previous to capital investment, or as Veblen puts it, "before American and German railway traffic was good for anything much," (1915) these other nations were in a good position to glean the British engineers' insights and incorporate these into their own budding rail systems.

In cases where the installed base cannot be overturned economists speak of the creation of a *path dependence*: technologies designed following crucial junctures (such as gauge width) thereafter come to be shaped by past technological decisions. For example, Veblen describes with amusement the UK's "restraining dead hand of their past achievements" (1915) embodied in the "silly little bobtailed carriages" which carried a much lower tonnage than later US and German rail cars, yet, he argues, no easy transition could be made as "the terminal facilities, tracks, shunting facilities, and all the ways and means of handling freight on the oldest and most complete of railways stems, are all adapted to the bobtailed car." (Veblen 1915). Path dependence is not solely a case of investment in technologies, it also a matter of persistent social organization and human training. Paul David (1984) has illustrated this through his study of the QWERTY configuration in keyboards. This configuration was historically developed to slow human typing, thus preventing mechanical keys from jamming – no longer a driving concern for some time but which today cannot be overturned because the arrangement of keys is embedded (a 'sunk cost') within not only hardware and software but also the typing skills of innumerable writers, clerical workers and programmers.

Returning to Veblen's mechanisms for latecomer advantage, he offers two explanations. The first and more obvious form is an obsolescence in the face of technical advance: a clear shift in particular technologies that require a replacement of the physical capital. In IT this is visible as continuous overturning of hardware, and software versioning. To the extent that the

---

“"These very elements of 'backwardness' in the logical and governmental aspects of legal development [in Germany] enabled business to produce a far greater wealth of practically useful legal devices than had been available under the more logical and technically more highly rationalized Roman law" (Weber, 1968 :II, 688, quoted in (Schneider 1971)).

inertia of the installed base cannot be overturned it is what Veblen called a 'depreciation through improvement.' Here the value of the technology does not simply disappear in the face of advance: its use continues 'through inertia' but it becomes incrementally less efficient or effective.

Along with depreciation through improvement, Veblen also identified a more subtle but pervasive obsolescence which is through 'an altered technological situation' (1915: 29-30). Rather than an individual technology being superseded, a systemic shift reconfigures all relations. Veblen illustrates this concept through a US steel corporation that became outdated not because of the technologies within the factory but because of transformations in the richness of raw materials and transportation methods of these materials. Means for producing cheaper and purer materials in the US had lowered the costs of transportation and had led to changes in the development pattern of the rail system; eventually the location of the steel corporation come to be outside the major transportation hubs and thus too distant from accessible labor pools. Such obsolesce through altered technological situation cannot be remedied by a replacement of equipment; the obsolescence is systemic. To a large extent Veblen sees such change as contingent and difficult to plan for, however it contributes to his thesis on advantages for a latecomer who can jump directly into a contemporary technological situation without bearing the costs of transitioning previous physical capital.

Similarly, the 'technological situation' of CI is being continuously reshaped, not simply through novel individual technologies but through reconfigurations of entire informational development strategies. Witness the availability of increasingly powerful computing cycles in personal computers; the dropping prices for data storage; the increasing bandwidth for networked communications. As with changes in the hubs for rail transportation, each of these transformations in information technologies has contributed to a redistribution of 'where computing happens,' 'where storage happens,' and the associated costs of collaborating across physical distances. While we know the importance of metadata to enable reuse and preservation, it is not yet clear what should be the content of that metadata: what do we need to capture to enable the applications of the future? Will a standardized metadata language emerge within particular disciplines (Millerand and Bowker in press) or even across the sciences? We are on the cusp of many decisions about the fundamental architecture of CI and many of these decisions cannot be made at the level of individual IT projects, even those at the scale of 'disciplinary infrastructures' such as GEON or WATERS). These come closer to institutional decisions about a future architecture of the technological situation.

5. Discussion: Three propositions

Even at this early stage in our understanding of CI development, we feel there are propositions that can serve as the impetus for improved design and development of infrastructure for e-social science (i.e., as contrasted with the experience in the natural sciences). These propositions flow from the consideration of the tensions in development of CI identified earlier in this paper, such as *project or facility*, *development or maintenance*, and *inclusion or innovation* (see also, Ribes and Finholt 2007). We present the following three propositions to guide initial application of 'lessons learned' from CI activity in the natural sciences.

Proposition 1: The stability of funding shapes the sustainability of the emerging infrastructure.

Short-term funding creates incentives for participants to develop incremental contributions to community resources, or, when more ambitious, deliverables in the form of proof-of-concepts, demos, or prototypes. This is because participants have shorter windows to show results between project reviews, but also because they must be thinking ahead to the next funding cycle. If we seek to develop sustainable infrastructure the institutions of e-social science (such as NSF) must adopt or invent mechanisms that permit long-term funding. For example, LTER is funded as an NSF program, a 'line item' which is reviewed once per decade; and WATERS is in a trajectory for funding under MREFC[11] (which from the planning stages to implementation span almost 20 years). Alternately, funding infrastructure under research grants (3-5yrs) could be ameliorated by creating clear venues for continuance of the project: possibilities for funding renewal, or project adoption by an institution of science.

Proposition 2: In order to support research, development and maintenance we must either change reward structures or incorporate new kinds of participants in CI efforts.

The sciences are structured to reward (e.g. recognition, tenure, funding) the production of new knowledge, new data or new analytic approaches. In the case of infrastructure, however, the span of kinds of work that are required and kinds of output that are desired is much broader. The work of technology development and deployment (or as Jane Fountain has described the process of making IT accessible for practical everyday use: enactment (2001)) demands considerations for collecting user requirements, 'hardening' systems, designing for usability, extensibility, backwards compatibility and so on. The sustainability of infrastructure requires continuous maintenance work, repair and upgrade. Yet it is precisely these activities of enactment and maintenance that fall outside the usual reward systems within the sciences (natural, engineering or social sciences, alike). Today, junior professors in paleobotany, systems engineering or sociology will not be well regarded by a tenure committee for having done the work of contributing their data to an accessible archive, nor for having worked long nights debugging and rebooting the computational resources that serve thousands of scientists in their work.

One often suggested solution includes transforming the reward structures of the sciences. For example, the proper registration of data within accessible databases could be rewarded on par with publication; the development of a visualization tool could be assessed as a novel methodological contribution; or the creation of a metadata standard could be viewed as a form of new knowledge. However, attempting to alter incentives is a slow and uphill undertaking. The reward systems of the sciences are deeply entrenched within their multiple academic institutions: editorial staff of journals, publishers, funding agencies, and departments. Another possibility, then, is the reconstitution of participants in the design of CI. While most eScience ventures are currently spearheaded, managed and enacted by scientists themselves (domain + computer) – responsibility could be distributed to others already rewarded for deployment and maintenance work. For example, LTER has a well organized subsection of information managers who maintain scientific databases and the systems for accessing them (Baker, Benson, Henshaw, Blodgett, Porter and Stafford 2000). LEAD, which did not originally receive funding for requirements collection, sought out additional resources and partners to fill these roles (Lawrence 2006; Lawrence, Finholt and Kim 2006). This strategy does not entail transforming 'career trajectories and reward systems,' instead new actors are added who provide services, perform upkeep and ensure the usability of systems.

---

11 Major Research Equipment and Facilities Construction. A special category of funding from NSF, particularly directed at developing large scale scientific instrumentation and sensor networks.

Proposition 3: There is a great danger of reproducing divisions between the resource rich and poor among social scientists -- such divisions will come to be reflected and sustained in technological development trajectories.

One of the strongest rationales for CI is its capacity to enable interdisciplinary research, crossing boundaries of method, technical language, or data structure. Yet in building such 'umbrella infrastructures' developers regularly encounter barriers to broad participation due to varying levels of readiness within and across disciplines. For example, Olson and Olson (2000) observe that successful adoption and use of CI demands both social and technical readiness. That is, a community must have a sufficient orientation to collaborate such that enhanced collaboration is perceived as a benefit. Also, where enhanced collaboration depends on technology, the community must have sufficient experience to distinguish useful systems and services.

A short-term consequence of differentiated development can be the marginalization of participants (such as described in GEON for paleobotanists and metamorphic petrologists); Star and Ruhleder (1994) have described these 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 described the phenomena of infrastructure 'capture' as resources accumulate amongst the 'haves' of science.

6. Conclusion: Simultaneously address the 'hard' and 'soft' foundations of infrastructure

It is a notably difficult and often unfruitful undertaking to form hard distinctions between the social and natural sciences. It is best not to ask how they differ, but rather to focus on how insights drawn from one can inform understandings of the other. Our research has explored three tensions framed by actors themselves as they attempt to develop long-term CI for the natural sciences. We speculate that learning from these experiences constitutes a form of latecomer advantage for social scientists as they plan for the development of e-social science. It should be noted that participants in the design of CI and scholars of CI have only begun to catalogue difficulties, strategies and 'best practices.' This paper can only be a beginning in the process of articulating the experience of building CI for the natural science as recommendations for e-social science. However, drawing from evolutionary economics and research in technology enactment we have provided rationale, and the outline of a method, for social scientists to 'mine' the recent history of scientific information infrastructure development in order not to repeat the mistakes of the past.

If there is a single recommendation to draw from our research, it is one that is regularly articulated with respect to cyberinfrastructure, but we suspect, will bear repeating as we shift into the development of e-social science. Specifically, the 'soft' foundations (i.e., social or human) of infrastructure will require just as much research and work as its 'hard' foundations (i.e., technical). This 'finding' has been repeatedly emphasized by social scientists collaborating with information technologists in building infrastructure (Ackerman 2000; Bowker, Star, Turner and Gasser 1997; Brown and Duguid 2000; Finholt 2004; Fountain 2001; Kling and McKim 2000; Ribes and Baker 2007; Star and Ruhleder 1994). To this commonly stated, but more difficult to follow, observation we would like to add the more subtle recommendation which mirrors Veblen's notion of the technological situation[12]: *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: 'socio-technical' (Hughes 1989) from the history of technology which encourage the analyst to simultaneously address organization, technology and context; or 'technoscience' (Latour 1987) from the sociology of science, which discourages distinctions between scientific practice and the materials which enable research.

***

Our future research follows two trajectories. Just as actors themselves have identified the challenges of transitioning from projects to facilities, they have also devised strategies to manage these tensions. First, we will seek to identify the strategies of the long-term that participants have adopted as a response: how have actors mobilized to transition projects to facilities? Following from this, we seek to identify the consequences of the long-term for the structure of CI projects: how has managing the tensions shaped the organization and technology within e-science endeavors?

Acknowledgments

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 Steve Jackson, Jeff MacKie-Mason and Paul Resnick who contributed in formulating Veblen's work. 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').

References

Ackerman, M. (2000) 'The Intellectual Challenge of CSCW: The Gap Between Social Requirements and Technical Feasibility', *Human Computer Interaction* 15: 179-203.

Baker, K.S., Benson, B, Henshaw, DL, Blodgett, D, Porter, J and Stafford, SG (2000) 'Evolution of a Multi-Site Network Information System: the LTER Information Management Paradigm', *BioScience* 50/11: 963-83.

Bowker, Geoffrey C., Star, S.L., Turner, William and Gasser, Les (eds) (1997) *Social Science, Technical Systems, and Cooperative Work: Beyond the Great Divide* (Mahwah, NJ: Lawrence Erlbaum Associates).

Bowker, Geoffrey C., Timmermans, Stefan and Star, S.L. (1997) 'Infrastructure and Organizational Transformation: Classifying Nurses' Work', in W. Orlikowski, G. Walsham, M. Jones and J. DeGross (eds), *Information Technology and Changes in Organizational Work* (London: Chapman and Hall).

[12] Thus even when speaking specifically of physical capital, Veblen was quick to note an implied assemblage of human training, organizational arrangements and technology: "the accumulated equipment, both material and immaterial, both in the way of mechanical appliances in hand and in the way of technological knowledge ingrained in the population"

Brown, John Seely and Duguid, Paul (2000) *The Social Life of Information* (Boston, MA: Harvard Business School Press).
Clarke, Adele (2005) *Situational analysis: grounded theory after the postmodern turn* (Thousand Oaks: SAGE Publications).
David, Paul (1984) 'Clio and the economics of QWERTY', *American Economic Review* 75: 332-37.
David, Paul A. (2004) '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*.
David, Paul A. and Spence, M (2003) '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*.
Edwards, Paul (2003) 'Infrastructure and Modernity: Force, Time, and Social Organization in the History of Sociotechnical Systems', in Thomas J. Misa, Philip Brey and Andrew Feenberg (eds), *Modernity and Technology* (Cambridge, MA: The MIT Press).
Edwards, Paul, Jackson, S.J., Bowker, Geoffrey C. and Knobel, Cory P. (2007) 'Understanding Infrastructure: Dynamics, Tension, and Design', *Report from "History & Theory of Infrastructure: Lessons for New Scientific Cyberinfrastructures"*.
England, J. Merton (1982) *A Patron for Pure Science: The National Science Foundation's Formative Years. 1945-57* (Washington, DC: National Science Foundation).
Finholt, Thomas A. (2004) 'Collaboratories', *Annual Review of Information Science and Technology* 36: 73-107.
Fountain, Jane E. (2001) *Building the Virtual State: Information Technology and Institutional Change* (Washington, D.C.: Brookings Institution Press).
Gerschenkron, Alexander (1962) *Economic backwardness in historical perspective* (Cambridge, MA: Belknap Press).
Glaser, B.G. (1978) *Theoretical sensitivity: advances in the methodology of grounded theory* (Mill Valley, CA: Sociology Press).
Hughes, Thomas P. (1989) 'The Evolution of Large Technological Systems', in Wiebe E. Bijker, Thomas P. Hughes and Trevor J. Pinch (eds), *The Social Construction of Technological Systems: New Directions in the Sociology and History of Technology* (Cambridge, Massachusetts and London, England: The MIT Press): 51-82.
Jewett, Tom and Kling, Rob (1991) 'The Dynamics of Computerization in a Social Science Research Team: A Case Study of Infrastructure, Strategies and Skills', *Social Science Computer Review* 9: 246-75.
Jirotka, Marina and Goguen, Joseph (1994) *Requirements Engineering: Social and Technical Issues* (New York: Academic Press).
King, Desmond (1998) 'The Politics of Social Research: Institutionalizing Public Funding Regimes in the United States and Britain', *B.J.Pol.S.* 28: 415-44.
Kleinman, D.L. and Solovey, M. (1995) 'Hot Science/Cold War: The National Science Foundation After World War II', *Radical History Review* 63: 110-39.
Kling, Rob and McKim, Geoffrey (2000) 'Not Just a Matter of TIme: Field Differences and the Shaping of Electronic Media in Supporting Scientific Communication', *Journal of the American Society for Information Science* 51/14: 1306-20.
Latour, Bruno (1987) *Science in action: how to follow scientists and engineers through society* (Cambridge, Mass.: Harvard University Press).
Lawrence, Katherine A. (2006) 'Walking the Tightrope: The Balancing Acts of a Large e-Research Project', *Computer Supported Cooperative Work* 15/4: 385-411.
Lawrence, Katherine A., Finholt, Thomas A. and Kim, Il-hwan (2006) 'Warm Fronts and High Pressure Systems: Overcoming Geographic Dispersion in a Meteorological

Cyberinfrastructrue Project', *Proceedings of the 39th Annual Hawaii International Conference on System Sciences (HICSS'06)*.
Lee, Charlotte P., Dourish, Paul and Mark, Gloria (2006) 'The human infrastructure of cyberinfrastructure', *Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work* NY, NY: 483-92.
Millerand, Florence and Bowker, Geoffrey C. (in press) 'Metadata Standards: Trajectories and Enactment in the Life of an Ontology', in S.L. Star and M. Lampland (eds), *Formalizing Practices: Reckoning with Standards, Numbers and Models in Science and Everyday Life*.
Monteiro, Eric (1997) 'Scaling Information Infrastructure: The Case of the Next-Generation IP in the Internet', *The Information Society* 14: 229-45.
Nentwich, M. (2003) *Cyberscience: Research in the Age of the Internet* (Vienna: Austrian Academy of Science Press).
Ribes, David and Baker, K.S. (2007) 'Modes of Social Science Engagement in Community Infrastructure Design', in Steinfield, Brian T. Pentland, M. Ackerman and N. Contractor (eds), *Proceedings of Third International Conference on Communities and Technology* (London: Springer): 107-30.
Ribes, David and Finholt, Thomas A. (2007) 'Tensions across the Scales:Planning Infrastructure for the Long-Term', *Proceedings of the 2007 international ACM SIGGROUP conference on Supporting group work, Sanibel Island, Florida, USA*.
Schneider, Louis (1971) 'Dialectic in Sociology', *American Sociological Review* 36/4: 667-78.
Star, S.L. (1999) 'The Ethnography of Infrastructure', *American Behavioral Scientist* 43: 377-91.
Star, S.L. and Ruhleder, K. (1994) '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* (Chapel Hill: ACM Press): 253-64.
Strauss, Anselm (1993) *Continual Permutations of Action* (New York: Aldine de Gruyter).
Trigg, Randall and Bodker, S. (1994) 'From Implementation to Design: Tailoring and the Emergence of Systematization in CSCW', in, *Proceedings of ACM 1994 Conference on Computer-Supported Cooperative Work* (New York: ACM Press).
Veblen, Torn (1915) *Imperial Germany and the Industrial Revolution* (New York, NY: Macmillan & Co.).
Walker, JD, Lehnert, KA, Hofmann, AW, Sarbas, B and Carlson, RW (2005) ' EarthChem: International Collaboration for Solid Earth Geochemistry in Geoinformatics', *Eos Trans. AGU Fall Meet Suppl.* 86/52.
Weber, Max (1978) *Economy and Society: an outline of interpretative sociology* (Berkeley, Ca: University of California Press).
Weedman, Judith (1998) 'The Structure of Incentive: Design and Client Roles in Application-Oriented Research', *Science, Technology & Human Values* 23/3: 315-45.
Womack, J.P., Jones, D.T. and Roos, D. (1990) *The Machine that Changed the World: The story of lean production* (New York: Harper Collins).