Ethnography of Scaling
Or, How to fit a national research infrastructure in the room

David Ribes
Communication, Culture & Technology (CCT)
Georgetown University
3520 Prospect St. NW, Suite 311
Washington, DC, 20057
dr273@georgetown.edu

ABSTRACT
Ethnographers have traditionally studied people in particular times and places. However, sociotechnical systems are often long-term enterprises, spanning the globe and serving vast communities. Drawing from three cases of research infrastructure development, this paper demonstrates a methodology in which the ethnographer examines scalar devices: actors’ techniques and technologies for knowing and managing large-scale enterprises. Such devices are enacted in and across concrete times and places; for the ethnographer they are observable as activities of scaling. By examining the enactment of scale we can better investigate diverse kinds of size and growth within sociotechnical systems.

Author Keywords
Ethnography; large-scale; methodology; grounded theory; actor-network theory; infrastructure

Categories and Subject Descriptors
J.4 [Social and Behavioral Sciences] Sociology
K.4.3 [Organizational Impacts] Computer-supported cooperative work

General Terms
Management, Design, Human Factors, Theory

INTRODUCTION
As an ethnographer who has dedicated his career to the investigation of “large” sociotechnical systems, such as research infrastructures, I am often asked (accompanied by a puzzled look): “how can you study large-scale things using ethnography?”

This paper outlines a methodological approach I have adopted throughout my ethnographic investigations, what I call ethnography of scaling. Rather than seeking to investigate the large-scale thing itself, the ethnographer asks of the actors: how do you know your enterprise? The key insight in this method is the recognition that anytime there is a “large” endeavor you will find actors tasked with managing the problems associated with its scale.

In an ethnography of scaling the ethnographer turns her or his analytic attention to the techniques and tools actors employ as they go about knowing and managing the scale of their enterprise: administrative actors such as project managers or principal investigators; actors trained in social sciences serving to map the size of such activities; or, technical actors who scale the capacity of servers, monitor bandwidth usage, or design architectures with an eye to extensibility.

All of these people, and many more, are concerned with representing and coping with the size of their endeavor. In order to understand how they manage size and scaling, the ethnographer need only follow in their footsteps. Whether “social” or “technical,” all of their work remains observable to the well-placed ethnographer in the form of practical activities, techniques and tools that I will call scalar devices.

Scholars within the ethnographic tradition have innovated additional approaches to investigate large-scale endeavors. For example, by traveling to a selection of targeted key sites, ethnographers have sought to “sample” geographically distributed enterprises, generating an understanding of the whole by assembling studies of its parts. Other approaches have sought to “scale-up” ethnography itself by adding more investigators to the research team. These are valuable and promising approaches that I have engaged in throughout my career, but this is not the approach I will describe in this paper. An ethnography of scaling turns these approaches on their head.

Rather than attempting to somehow match the scale of the object of investigation, in ethnography of scaling the investigator focuses on the work of actors as they go about knowing and managing that large-scale object. Observing scaling activities does not require a radical new approach.
Instead, it demands the sustained application of an ethnographic sensibility that takes scaling as its object of investigation. We need never change the framework or the mode of analysis, regardless of what we, or our actors, consider to be the scale of the enterprise. All the activities of scaling are observable on what sociologist Bruno Latour has called a “flatland” [25].

In order to elucidate this methodology, I will draw from three past ethnographic studies of the development of research infrastructure: GEON, WATERS, the Open Science Grid (OSG). Research infrastructures are sociotechnical assemblies of information technologies, resources and services that support the doing of science, today often called Cyberinfrastructure or eResearch [36]. These endeavors are large in many ways [8]; they are geographically distributed across nations or the globe; they are ambitious, promising to change how science is done; they are significant financial investments; they target many users from many disciplinary backgrounds; they develop national or international computational networks; and so on.

I will focus on three scalar devices I have observed that have been present in all my investigations of research infrastructure: the All-Hands Meeting, surveys and descriptive statistics, and benchmark metrics – each of these scaling activities seek to represent and manage the size and growth of these sociotechnical systems.

Beyond a methodological contribution to ethnography, this paper is also a substantive contribution to understanding infrastructure development, and sociotechnical systems more generally. Such enterprises bristle with scalar devices: surveys, requirements collections, social network analyses, benchmarks, metrics, evaluations and “meetings all the way down.” Developers are never satiated with one, two or even three results of such operations: they are forever seeking more, and new, handles on the scale of their enterprises. In the discussion section I will illustrate how multiple scalar devices are gathered together by actors in order to stich together representations of their endeavors.

ETHNOGRAPHY: Copresence, Artifacts and Practices of Representation

Ethnography is a long and varied tradition. In its most basic form, the observer is a participant in the activities of interest. By observing and participating, the ethnographer can recount, theorize or intervene on a rich array of possible themes of interest such as shared meanings and diverging understandings; practices and artifacts; collaborations and conflicts, and many more.

Rather than traversing a potted history of ethnography – from its roots in anthropology and sociology to its current use in infrastructure development [38] – I will jump straight to the approaches that inform this methodology:

i- those that emphasize the production of copresence rather than requiring colocation;

ii- in which archival or trace evidence can serve as ethnographic data; and,

iii- that focus on material artifacts and practices of representation.

Traditionally, for the ethnographer, investigation occurs at a common site that is shared with the actors, enabling a rich interactional environment for participant observation. However, it is easy to recognize that many interactions occur in a manner that is neither collocated nor synchronous; the internet is the current exemplar of distributed, asynchronous activity, but there are many others, such as love affairs by correspondence or business meetings held on conference phones. Anne Beaulieu has offered the concept of copresence as a corrective to the emphasis on colocation in ethnography. The concept “opens up the possibility that co-presence might be established through a variety of modes, physical co-location being one among others.” [1]. In short, her argument is that today we achieve copresence in multifarious ways: by commenting on the same news feeds as our colleagues, through the “thin data” of an instant message, as well as by sharing a cup of tea at a café. Citing sociologist Erving Goffman, Beaulieu notes that “physical presence is not equivalent to availability for interaction.”[1]. The ethnographer can generate a sense of copresence by listening in on a conference call, by joining a feed, by reading conversations on a message board or reconstructing activities from a detailed archive.

Ethnography need not be “live,” it can be based on data generated in the past. In her study of the Challenger Launch Decision, Diane Vaughan [41] used the term historical ethnography to describe her practice of delving into the wealth of secondary data that NASA itself had produced through its investigation of the space shuttle explosion. What made her approach ethnographic, rather than historical alone, was her sensitivity to ongoing interaction (and copresence) in NASA.

My own investigations of Wikipedia [11, 12] and the Open Science Grid [35] have been based on archival traces documented in logs and in ticketing systems. In contrast to the thick description of classical ethnography [10], such traces are thin data. However, because the actors themselves in these online and distributed projects only have these same traces to make sense of distanced activities, the ethnographer and the actors are in symmetrical positions. In other words, in these two cases, actors and ethnographers both relied on thin data to go about their “thick activities” – though for different purposes.

Finally, material artifacts are important. In observing activities of scaling, practices of representing and representations themselves are central. “Large objects” are made available to interaction and practice by representing
them at a human scale. For example, in a study of architectural practice, Albena Yaneva [46] focused on how designers work on scaled physical models of a building. Yaneva describes the use of a design tool called a “modelscope,” a miniature periscope that is inserted into the architectural model, generating a viewpoint within its model space. A small model can thus be momentarily inhabited at the human scale, informing large design.

Different kinds of scaling activities are accompanied by their associated practices and tools: some of these tools are common across many spheres (e.g., the summative methods of descriptive statistics) and some are endogenous to only a few domains, such as the modelscope in architecture.

I will call an assembly of techniques, tools and representational conventions that are used to know and manage scale, a scalar device. In the empirical portions of this paper I will explore three specific scalar devices that I have seen used in all infrastructure development projects: meetings and agendas; surveys and descriptive statistics; and computational metrics and benchmarks.

A note on terms

I will use the term observable to refer to that which is visible (or audible, as with a conference call) to the ethnographer, whether through unassisted observation or through the use of digital traces, fieldnotes, audio and video recordings or other documentary materials. I will use the term actors to refer to those people the ethnographer is observing: they are actors in the sense that they are engaged in the enactment [43] of the scale or growth of their enterprise. My empirical topic is research infrastructure and so I will also call the actors infrastructure developers. I will use the term object of investigation, or object, to denote the target of scaling activities. Objects can be large or small (e.g., a group, an institution, a global network) as defined and worked-over by actors, but in a strict ethnography of scaling, large objects are only available as that which can be inspected in interaction as talk, practice, and inscription.

Large-Scale Objects in Ethnography

In the brief literature review that follows, I distinguish three approaches to ethnography of large objects: sampling, scaling-up ethnography and ethnography of scaling. I demarcate these approaches for heuristic reasons but in practice I have always used them together: I have always travelled to the distributed sites of my actors, and I have always had collaborators in the study of infrastructure, whether advisors, colleagues or research assistants. I focus on ethnography of scaling here, in order to add it to our repertoire of techniques.

Sampling in Ethnography

A popular approach to studying large-scale objects has been what anthropologist George Marcus has called multi-sited ethnography [29], but which in this paper I will refer to as sampling in ethnography. In studying a global scale object, such as colonialism, market regimes, state formation, or nation-building, the ethnographic investigator “moves out from the single sites and local situations of conventional ethnographic research designs to examine the circulation of cultural meanings, objects, and identities in diffuse time-space”[29]. Thus, in order to investigate, say, a globalized restaurant corporation, an ethnographer may travel across nations to synthesize, contrast and compare that restaurant’s venues. Or, rather than literally traveling to sites, the investigator may use documentation to trace the movement of, for example, hamburger meat as it travels from farm, to processing plant, to restaurant, in order to arrive on a plate. As anthropologist Ulf Hannerz notes, “multi-site ethnography almost always entails a selection of sites from among those many which could potentially be included.” [14].

One of the earliest ethnographic sampling approaches for the study of large sociotechnical systems is rooted in CSCW: Susan Leigh Star and Karen Ruhleder’s classic study of the Worm Community System (WCS) infrastructure [39]. With a healthy travel budget in hand, Star [38] described flying from one site of implementation to another. At each site her ethnography would, in part, begin anew as she got to know the people, their scientific interests, the arrangement of their lab, and how (if at all) they were using the WCS.

The advantage of this method is its scope. At each site it is possible to observe, first hand, practices and discourse, material arrangements, or conflicts and cooperation. The “large,” or “global,” qualities of the object become specific by observing concrete connections and settings. As Marcus put it, in this approach “the global is an emergent dimension of arguing about the connection among sites” [29].

The disadvantages are: potentially thin ethnographies, cost, time, exhaustion of the investigator, and a healthy dose of ongoing awkwardness as each site poses new relations for the ethnographer. In ethnography the investigator must demonstrate that she or he has spent enough time engaged in interaction (collocated or not) to be able to faithfully relate actors’ categories or practices. The grounded theory tradition that informed Star’s work focuses on theoretical sampling (rather than representative sampling as with statistical methods [13]) in which it is important to return to the field site and target the elaboration of a concept: i.e., has sufficient sampling occurred to saturate the emerging theory? In any method that relies on sampling, saturation is a challenge.

Recently, scholars have displayed an interest in combining ethnography with other methods so as to better investigate large-scale objects. For example, there has been a recent spurt of papers using network analytic approaches in combination with ethnography [7, 18]. In such approaches,
the network analyst assembles relational traces (such as co-authorship or citation data) to produce an image of the object; then, by identifying certain focal nodes an ethnographer can be dispatched to investigate this site in richer detail, i.e., sampling.

**Scaling-Up Ethnography**

In a second style, the approach is to scale-up the ethnography so as to match the scale of the object of investigation. This means assembling a team of investigators who will tackle the same object. If the object of investigation is geographically distributed, then the team may also be.

Williams and Pollock’s study of enterprise resource planning software is an excellent example of a scaled-up study in which they investigated their object for extended periods, and in different settings of development and use [44].

The advantages are numerous. Having multiple eyes on different facets of the same object is invaluable. A classic danger in ethnography is providing an excessively unified view of the object (“a culture,” “a community,” or “shared meanings”) [9]. Multiple collaborators engaged in participant observation can better ensure a complex understanding of contradictions or competing definitions.

The disadvantages of scaling-up ethnography are shared with all forms of team-based and/or geographically distributed research [17, 21]. Coordinating the work of multiple colleagues is a challenge of its own, including working to harmonize the data collection and analysis of the team. Because ethnography is a methodology that can be associated with highly heterogeneous theoretical commitments, working with multiple collaborators can make investigating common objects challenging.

**Ethnography of Scaling**

Ethnography of scaling inverts the above approaches to large-scale objects. Rather than reconstructing the object by sampling it, or by scaling ethnography to match it, the key methodological recognition is that knowing and managing the size of the enterprise is a concern for the actors. The object of analysis for the ethnographer then becomes the methods, techniques and technologies used by actors to know and manage their enterprise.

This approach is similar to past ethnographic analyses focused on centers of coordination or calculation. For example, classic CSCW studies have focused on crisis and airplane control centers [15, 40]. Work-practice studies of centers of coordination are often engaged with the question: how do members practically manage the complexity and diversity of their enterprise? More recently, Karin Knorr-Cetina has examined global financial markets as they occur on the computer screens of distributed trading room floors [23]. Investigators execute these studies by turning their ethnographic gaze to those concrete places and times where such activities are occurring.

However, in order to study scaling we must also track-back, or “open the black box,” of scalar indicators themselves. For instance, in examining scaling, when the investigator is confronted with a statistic indicating the size of the object, the goal of the ethnographer is to uncover how this statistic was generated. An indicator of scale often hides the enormous organizational and technical work to produce it.

For example, the statement "we have 2341 members in our organization," renders invisible the activities that went into compiling this number. If this was done through a survey, then literally dozens of collaborators and hundreds of hours of work may have gone into the design, implementation, dissemination, harrying of respondents and then collating, analyzing and visualizing the resulting data to produce that single number.

An ethnography of scaling does not end at the indicator itself, it also tracks the course of indicators as they are deployed in managing the enterprise. The trajectory of the indicator “2341 members” does not end with its presentation as a finding, it goes on to be enacted in future organizational activities: the indicator may be used to decide funding allocations, to identify a needed growth in the organization, or to estimate server capacity required to support a website.

Ethnography of scaling sits at the intersection of the ethnomethodological insistence on the inspection of concrete circumstances [16] and the actor-network sensibility to trace concrete trajectories that weave across circumstances [6]. In broad stroke, ethnography of scaling involves a three part investigative focus on:

i- the development and deployment of a scalar device;

ii- the resulting indicator and its reception; and

iii- the down-stream consequences of indicators as they are wrapped into organizational action or design.

It is across the trajectory of these activities that the scale of an enterprise comes to be represented, known and then managed.

A key methodological prescription in this approach is maintaining a **scalar agnosticism**. The investigator must bracket any preconception of the size of the object [6]. Whether the object is large or small (or something else – see the Discussion) is a concern for the actors. It is they who are engaged in knowing and managing scale, not the ethnographer. The concern for the ethnographer is being at the right place, at the right time, to observe activities of scaling, or, to gain access to the trace materials of those activities.
For example, in the statement “our website gets 12,485 hits a day,” the ethnographer should consider the methodological difficulties that actors encountered in generating such a distinct count. What debates went into determining hits to a website: Do crawlers and bots count? What about multiple visits from the same browser? Should organic search arrivals (by a search engine) be treated differently than links from websites? This work of debating what does and does not count [30] is all but erased in the final indicator. In an ethnography of scaling these must be recovered in order to understand how the “size” of the website (its hits per day) came to be known.

Most difficult is that many scaling activities are fully automated and algorithmic. Unpacking how such figures are produced may require a kind of code spelunking that is outside the traditional skillset of the ethnographer. This is an example of infrastructural inversion [5], where the ethnographic investigator must take a “black-boxed” indicator of scale and actively seek to “open it up” by examining the, often highly technical, work that went into its production. The third empirical example in this paper (the Open Science Grid), demonstrates this spelunking by focusing on the actors that innovate and revise such automated scalar devices.

**REFLECTIONS ON FIELDWORK IN GEON**

I initially adopted the methodology described in this paper while doing the fieldwork for my Ph.D. dissertation [32]. For my thesis I investigated GEON: the geosciences network. GEON was a very ambitious endeavor, with the goal of implementing an umbrella organization drawing together informational resources in the broader earth sciences e.g., geology, geophysics, paleobotany and many other geoscience disciplines [22]. The mandate for this project was to link the scientific data, mapping tools and other geoscience disciplines [22]. The hope was that such a meeting would help in formulating a unified vision across their heterogeneous activities.

The concern in large endeavors is, for example, that the social ties of members will dissipate, that geographic distribution will be an impediment to collaboration, or that a lack of coordination will lead to duplicate efforts. In short, running a large sociotechnical organization is hard and those who are doing so know it.

Meetings are a common, and mundane, method that actors use to ameliorate such problems [4]. Meetings are gatherings that (re)familiarize participants with each other and with their heterogeneous work. In a distributed organization, they are an opportunity to recount recent accomplishments to far off colleagues. In a multidisciplinary enterprise, explaining scientific topics and methods face-to-face can help members overcome the balkanization of technical specialization. In GEON, the technical topics included: earth science, its various objects and methods of investigation, as well as the promised information technologies of cyberinfrastructure, such as computational ontologies and grid computing – none of these are self-explanatory.

GEON had twelve principal investigators (PIs) at eight universities across the U.S. Each of these PIs had a local team at their institution working on aspects of the project: over one hundred computer scientists, information technologists, earth scientists, and administrators regularly participated in GEON. The All-Hands Meetings were one occasion that would bring a large proportion of these specialists together (but never truly all). Meetings served followed by meetings. Notably, scaling activities need not be esoteric: they are often mundane. In the section after, I turn to the more technical practices of generating descriptive statistics.

**“GEON, like the turtles, is meetings all the way down”**

The organizational core of GEON was at the San Diego Supercomputer Center (SDSC) on the University of California San Diego campus. The sociology department was in the next building over, and so, to conduct fieldwork I needed only to cross a slim walking path. While names like GEON and Supercomputer Center may evoke grand imagery of towering servers and vast computing resources, the majority of this fieldwork was conducted in one of two rooms: an auditorium that never housed more than sixty participants, and a small conference room on the third floor.

Twice a year, protected from the racket of computers and servers by the fabric-covered walls of the SDSC auditorium, the GEON members engaged in a genre of meeting I saw enacted again and again, and thereafter in many other cyberinfrastructure projects: the All-Hands Meeting. These gatherings sought to bring together “all members” of the enterprise, such as geoscientists, computer scientists, and educational specialists. The hope was that such a meeting would help in formulating a unified vision across their heterogeneous activities.

In this first case, I focus on how actors engaged in scaling by using the well-worn activity of meetings and taking of minutes; these meetings were preceded by meetings and

---

1 Now defunct, the GEON portal can still be accessed at geongrid.org.

2 In the later years of this investigation my study was wrapped into a comparative investigation of cyberinfrastructure – see interoperability.ucsd.edu — which used sampling methods and scaled-up the ethnography.
to render visible the distributed activity of participants as a local observable performance.

Meetings are like Russian dolls, nested. In advance of an All-Hands Meeting, PIs and administrative staff at the SDSC would arrange other meetings to write an agenda and settle the order of presenters (and to be sure that lunch and coffee would arrive on time!).

An agenda is a mundane little model for how a meeting will go. They are scalar devices. Without an agenda, or an equivalent of the work it does, attendees at the All-Hands Meetings would not know when they were presenting, and would constantly badger their hosts about coffee breaks and lunch.

The order of presenters on an agenda is meaningful, and it can even be touchy. Usually each of GEON’s two lead-PIs were at the top of the agenda, inaugurating the day – one from geoscience and one from computer science, for balance. Then came the National Science Foundation (NSF) program officers, speaking in the name of the agency that had funded GEON. Then followed brief presentations by the other PIs and Co-PIs, and, occasionally, a young ethnographer to speak about community and interdisciplinarity [33]. Agendas map out these arrangements in advance, making the All-Hands Meetings easier to navigate by adding a little bit of order.

In a single morning of presentations a keynote speaker from NSF would outline the agency’s plans to build a national cyberinfrastructure; next, in a ten-minute presentation, a programmer would outline his detailed work over the past six months of capturing earth science knowledge in an ontology; and, then, a geoscientist would report her work on the crustal structure of the Western U.S. These presentations sought to summarize the organization of GEON, its environment or technical advances, bringing everyone “onto the same page.”

While GEON sought to implement high-end computing resources and experimental technologies, its meetings were no more high-tech than any contemporary conference presentation. At the first All-Hands, in 2002, one absent principal investigator did telecommute and showed up projected on-screen; but, technical difficulties delayed his appearance and eventually poor sound quality forced the reappearance of a rather old-fashioned looking conference phone, placed discretely in a corner of the room within reach of a jack. This little story should surprise no one who studies or does distributed work.

Throughout such meetings multiple participants were capturing the details of the meeting in notes and audio recordings:

Who’s doing what? I always write everything down, if I don’t I’ll never remember what to say when someone asks me what we are up to in GEON (PI, interview).

I attended the first meeting of GEON in 2002. At the end of my research I attended my last meeting in 2006. That first day, before they had built anything, I was able to look across ten meters (rather than the country) and see “all of GEON” in its days of simple composition: geoscientists, laptops, computer scientists and coffee. In the intervening time GEON members developed a plethora of information technologies for the earth sciences and built out its collaborations with other institutions: computational ontologies, integrative mapping tools, an affiliation with NASA, and innumerable computing cycles donated by the TeraGrid. But when I looked across the room during the 2006 All-Hands Meeting, three years later and after all these developments, even then I still saw geoscientists, computer scientists, and slides projected on-screen.

For the ethnographer watching presentations at a meeting, these were all observable as activities of scaling. Bruno Latour, in his studies of scientific laboratories and representations, reminds us that whether the object of investigation is a microscopic bacterium or a vast stellar constellation, in the end they must all be rendered available to the manipulation of scientists’ eyes and hands as print-outs or images on a screen [24]. Similarly, in the work of building national, interdisciplinary infrastructures, meetings are the laboratories of their size: agendas, slides and notes are their scalar devices helping to fit the geosciences and their future information technologies into the room for all to see.
THE LITTLE TECHNOLOGIES OF COMMUNITY

Ultimately, the developers of a large infrastructure, such as GEON, still number in the scores or, at most, hundreds. But the target communities of such enterprises – the future users of an infrastructure – are an order of magnitude larger. To know and manage their size, different scalar devices are needed.

This section explores the more esoteric scalar device of surveys. Surveys seek to investigate either an existing community or the future users of an infrastructure, in order to inform the design of that infrastructure. Surveys are “esoteric” in the sense that there is an art and a science involved in their design that seeks to reveal the phenomena of, for example, “community opinion” or “user requirements.” Often, the technical work of enacting a survey is delegated to those trained in social scientific methods who create and administer the survey, as well as analyze and represent findings. In an investigation of scaling practices, the analytic gaze of the ethnographer turns toward social scientists and their techniques. Such is the case with WATERS, which employed social scientists to help them make sense of a sudden growth and increased diversity in their targeted users.

The recent visions of research infrastructure, such as cyberinfrastructure, has targeted very large user bases: “Rather than supporting teams or groups, [developers] speak in terms of communities, disciplines and domains.” [36]. A successful research infrastructure must support the activities of these broad and diverse constituencies. But how to know what these users want?

It is rarely possible to have “all of the future users” of a domain infrastructure in a single room, as it was possible to have “all the developers” in the rooms of GEON. In the U.S., even a small domain of practicing scientists will number in the thousands – not to mention their technicians, information managers, and other staff so crucial to the functioning of infrastructure [28]. Even if such a vast gathering were possible, the cacophony of voices that would ensue would help no one resolve the issue of how to build better information technologies. But a crafted survey can target just those issues of interest to developers and then summarize findings in tidy bar charts and pie graphs. In doing so, a survey, by its design, will shape findings, intentionally or unintentionally, and may also altogether leave out topics that are of less interest to developers.

Nicholas Rose has called the survey, along with many other such scalar devices, the “little technologies of community”:

Over the second half of the twentieth century, a whole array of little devices and techniques have been invented to make communities real. Surveys of attitudes and values, market research, opinion polls, focus groups, citizens’ juries and much more have mapped out these new spaces of culture, brought these values and virtues into visibility and injected them into the deliberations of authorities. [37]

The survey is perhaps the most commonly employed “little technology of community” within the development of large sociotechnical systems. All the research infrastructures I have investigated have at some point commissioned a survey of their own membership or constituencies. Because within systems design it is often considered that “the community should drive development,” surveys can serve to demonstrate that the community is being given voice in design decisions. The size, complexity and diversity of community – a multitude of heterogeneous users – can be accounted and summarized through descriptive statistics.

Surveys are tools of governance in that they empower a small group of people to speak in the name of a larger one. In some cases, a survey successfully assuages fears by assuring users that their interests are represented in the design of the future infrastructure; in other cases, surveys infuriate, as users feel silenced, underrepresented or misrepresented [19]. This is one feature of interest to the ethnographer of scaling. This should also be of the utmost importance to those developing infrastructure and to the constituency an infrastructure intends to serve, because, as with any scalar device, indicators will play a role in downstream design [45]. In the case of WATERS, survey findings were used to allocate resources in the planned development of the infrastructure.

The WATERS Event

WATERS, like GEON, had ambitious goals to develop cyberinfrastructure, but targeted the disciplines concerned with the environments of water: rivers, rainfall, sewers and estuaries. Unfortunately, it was inaugurated with much more fractious beginnings than GEON. WATERS was born out of the top-down matrimony of two cyberinfrastructures for the domains of hydrology and environmental engineering3. Because both research fields seek to understand the environments of water, officers at the National Science Foundation decreed that a single infrastructure would avoid unnecessary duplication of resources. Through WATERS, both fields would develop a single interoperable set of observatories and information resources for the unified domain. I will not further recount how this merger came to be, which in past writings I have called “the WATERS event” [34]. Instead I will focus on the consequences and responses of participants.

---

3 WATERS was the Water and Environmental Research Systems Network. It sought to bring together previous efforts in hydrology led by CUAHSI (Consortium of Universities for the Advancement of Hydrologic Science) and environmental engineering’s CLEANER (Collaborative Large-Scale Engineering Analysis Network for Environmental Research). WATERS and CLEANER are now defunct, but CUAHSI continues its information systems efforts on several fronts.
The initial reaction by both groups of researchers to the merger was heated – polemic even. It should surprise no one that the WATERS event was a controversy: there was a great deal at stake for all parties. Both sides felt that the rug had been yanked from beneath their vision of research infrastructure. At the heart of the matter was the question of resource distribution: what disciplinary groups would get funding? From the forthcoming observatories, what institutions would hold new troves of rich data? Most contentiously, what would those data be about? And thus, at the very kernel of designing research infrastructure: what disciplinary objects of investigation would be made easily researchable and which would receive less, or no attention?

From the perspective of the leadership of this newly merged cyberinfrastructure, WATERS was clearly a much larger endeavor than the two initial projects individually: but what kind of large? And how would it affect plans to develop research infrastructure? For an ethnographer who was well positioned to observe the action, the WATERS event led to a self-conscious and observable activity of scaling.

Questions about scaling-up in WATERS were of three kinds:

i. **Scaling-up users:** WATERS would serve two domain communities. Did this mean doubling the user base? Probably not, since it was known that hydrologists and environmental engineers overlap in many ways. So then, how big was the future user community of WATERS?

ii. **Scaling-up of objects of investigation:** A common phrase that emerged within WATERS was that “environmental engineers care about water quality while hydrologists care about water quantity” or, in other words, engineers are concerned with, for example, whether water can be used to safely irrigate crops, whereas hydrologists care about how fast water will erode a riverbank based on its flow. What this little shorthand indicated was that these communities are concerned with different objects of research. A single research infrastructure would have to support both, and, thus, would have to collect more kinds of data, would have to develop more instruments to collect those data, and hire more information managers familiar with those data. But how much more data, and what kinds? These specifics remained an open question.

iii. **Scaling-up of the institutional resources to be integrated:** Environmental engineers and hydrologists are already enmeshed within their existing funding agencies, scientific societies, and disciplines. They have vast extant data archives and distinct software they use to do their work. The promise of cyberinfrastructure is to interoperate databases and support sharing of software tools. Would WATERS mean a doubling of such resources to be integrated? Again, probably not, since the communities overlap; but again, how much?

In order to make sense of these diverse forms of growth, and to help clarify the fuzzy boundaries between environmental engineering and hydrology, a key strategy in WATERS was to commission surveys of these communities.

I was fortunate that as I began to study WATERS, it was my new post-doctoral advisor, Thomas Finholt, and his previous postdoc, Katherine Lawrence, who had recently conducted one of these surveys [27]. Once again, I did not need to travel any further than up the hall at the University of Michigan iSchool, in order to observe these scaling activities. From my inquiries, I garnered all kinds of informal details about how the survey had been conducted, such as the “incentivization” of participation in the survey by including a two-dollar bill in the envelope with the questionnaire. However, the basic methodology of the survey was *pro forma*: Lawrence and Finholt sent a questionnaire to three hundred randomly selected people from each of two scientific societies that stood-in for the two domain communities of WATERS [27].

The survey revealed a landscape of relatively clear differences and commonalities between and across these two communities, defining the boundaries of each field while also pointing to promising avenues for developing a common research infrastructure. As had been expected, environmental engineering and hydrology overlapped in many respects, while diverging in others:

> each group has very different needs for data sources. For example, [hydrologists] are substantially more reliant on streamflow data [while] engineers had a substantially greater interest [...] in the USGS and EPA water quality databases. [27]

These findings matched what was already known about these communities (i.e., quality vs. quantity), but now in the precise numbers of descriptive statistics and by identifying specific databases. Commonalities also emerged. For example, to the horror of developers, but to the surprise of no one, the most commonly used data storage format for all groups was the Excel spreadsheet.
Surveys reveal disciplinary difference as ordered preference and heterogeneous practice but they also point to shared goals, widespread objects of interest and commonly desired computing resources. The results of the survey translated the future users of WATERS into the precise language of descriptive statistics and rendered findings in colorful and easily readable charts. The boundaries of engineering and hydrology, once intuited but ephemeral, were buttressed by such findings, constituting a detailed statistical reality of what was formerly impressionistic disciplinary difference.

The trajectory of these findings did not conclude with the publication of the survey. This is merely a mid-point in the life of these new indicators. As I continued my investigations, these charts and graphs would reappear continuously, projected on screens as I attended the meetings of WATERS. Figure 1 is a photograph (taken by me) of a slide that summarizes the cultural qualities of hydrology as a discipline, i.e., as the slide claims: hydrologists have a “lower emphasis on data ownership” and are an “integrative discipline.”

Techniques such as surveying do not automatically compel users or developers to believe their findings. In principle, whether findings are representative of a community can be a matter of debate. The objectivity or competence of the survey can come under fire. However, I have never observed such a debate questioning a survey itself (though, see the outcome of the OSG survey in the next section).

What I have observed are debates over whether survey findings have been properly incorporated into the design of the infrastructure. For example, at the same meeting as the slide in Figure 1 was displayed, during the question period, a self-identified hydrologist stood up and made the statement:

We are all here deciding what WATERS will look like. And from what I see it’s all quality and no quantity. And hydrology just isn’t going to stand behind that. (field notes)

Referring to the shorthand I described above – that hydrology is concerned with quantity of water – the scientist argued that the interests of his discipline, in the form of plans to develop observatories and sensors, had not been properly reflected in the design for WATERS he was hearing discussed at that meeting.

A great deal of ink has been spilt demonstrating how surveys produce findings that are tidier than the sentiments of their respondents [20]: opinions held by individuals in shades of grey are captured for the aggregate in black and white. This is certainly the case. But we should also keep in mind that this is also the strength of the technique: the survey is a scalar device to reduce the cacophony of the crowd to a manageable diversity of voices. The full range of dissent and eccentricities in the domain are lost: Laura, the environmental engineer who insists on recording her data on paper; Bob, the hydrologist who checks for strychnine in river water by counting the average number of eggs in a bird’s nest. Descriptive statistics do not capture such individual practices very well – granularity is lost. What is gained from a survey is a representation of community rendered as discrete groups and well-defined stakeholders. Each scalar device only preserves particular relations as it redescribes an object in its distinct representational form.

**SCALABILITY: A design virtue in the metric**

The two examples above could be characterized as cases of organizational scaling – they are about the number and qualities of people involved in projects. Lest the misconception be allowed to stand that scaling is a matter of the social, narrowly construed as only about humans, this last section will focus on a technical outcome: scaling computing capacity. This is perhaps the most common use of the term “scaling-up” within cyberinfrastructure development. Correspondingly, one of the greatest virtues in the design of infrastructure, and information systems more generally, is that they be scalable (can support growth), or the related virtues of extensibility (can expand into new capacities) and modularity (the use of common protocols that make discrete components interchangeable). Scalability, extensibility and modularity are the triumvirate gold standards of good design for today’s technical architecture development efforts concerned with future growth. These too can be observed as a matter of scaling activities as long as the ethnographer is willing to spelunk into code, algorithms and esoteric metrics.

Here I will briefly explore activities within the Open Science Grid (OSG) as they planned to support the reopening of the Large-Hadron Collider (LHC) in 2009,
which would demand a significant scaling-up of the computing power they were offering\(^4\). The kernel of the OSG is to make available processing power “on tap” to geographically distributed scientists. In sweeping (and not altogether accurate) terms, the OSG is an instantiation of the infrastructure-as-a-service (IAAS) model of cloud computing. *Figure 2* shows a chart of usage within OSG, and its projected future growth. How was such a projection made and what was concluded from it?

In order to know its scale OSG must be instrumented with tools specific to computing. We are used to seeing metrics of computing use, and we tend to ignore or take these for granted — e.g., megahertz, flops. But within high-performance computing enterprises there are technical experts dedicated to developing and deploying new metrics and tools to keep track of usage — something that is more challenging when many computing cores are distributed around the globe working in concert.

I track the activities of one technical specialist, Brian Bockelman. Trained in computer science with a concentration in metrics, Bockelman has worked for many years to, as he puts it, “give estimates of the “size” of the Open Science Grid and demonstrate why such a number is hard to come by.” (quotes in original)\(^2\).

To measure the “size” of OSG, Bockelman et al.\(^3\) first attempted the administration of a survey, but found responses were riddled with “human errors” as local administrators failed to supply the correct metrics. More importantly, the results of the survey were static even as new sites were quickly being added to OSG: “survey results would not scale as the grid grows.”\(^2\)

Instead, in the OSG they have relied on a piece of software called Gratia\(^3\) that is used for job accounting. A job is a unit of work for an operating system. By installing Gratia at each OSG site they were able to track jobs: the OSG averaged between 5 and 11 million jobs per month between May 2008 and March 2009\(^2\). Even this metric is crude, as it treats all jobs as though they have the same computational cost. They do not. And so Bockelman combined Gratia with a metric for aggregate computation: Wall hours; “if a job ran for one hour on one core consuming 50% the CPU, we say it has .5 CPU hours or 1 Wall hour.” With this metric in hand, Bockelman was then able to extrapolate the growth of the OSG “using a linear fit based on the last 10 months, [Figure 2] shows the projection for the following 6 months”\(^2\).

Based on his projection, Bockelman concluded “This burst of growth indicates the OSG is in a good position for delivering further growth when the LHC collisions start,” but also added the caveat that growth is linear only over the academic year; during the summer, it stalls. He speculated: “It appears that the primary variable in grid usage is still sociological,”\(^2\).

Bockelman is neither a PI nor a project administrator. It is not only managerial actors who are concerned with scaling; we will encounter the deployment of scalar devices throughout sociotechnical systems and for many purposes. Even within the “depths of the machine” – and as with any scalar device – metrics and tools are developed to know and manage the growth the object. These indicators serve to inform actors about the future of OSG as they organize and plan to selectively allocate computing power where it is needed and to add more resources in advance of shortages.

**DISCUSSION: What is to be done with large objects?**

**Proliferate scalar devices**

One day the husband of the woman is called to the artist’s studio. “What do you think?” asked Picasso, indicating the nearly finished picture. “Well…” said the husband, trying to be polite, “it isn’t how she really looks.” “Oh,” said the artist, “and how does she really look?” The husband decided not to be intimidated. “Like this!” said he, producing a photograph from his wallet. Picasso studied the photograph. “Mmm…” he said, “small, isn’t she?”

— Gareth Morgan, Images of Organization\(^31\)

Allow me to recapitulate the concept of *scalar agnosticism*. At the beginning of this paper I defined it as

---

\(^4\) opensciencegrid.org In the past, I have investigated the delegation of coordinating work to software systems for cycle sharing in OSG\(^35\).
Scaling has always also been temporal, whether this has been a meeting agenda that structures the activities of an upcoming get-together; knowing the prospective users of a community infrastructure; or predicting the cycles that will be needed in order to compute forthcoming particle collisions. Part of the goal in activities of scaling is to manage growth, and part of growth is what is yet to come. Scaling is a future oriented activity.

Each of these ambitions for growth present unique kinds of problems to know and manage the scale of an enterprise, and each has distinct techniques and technologies to manage that form of scale. In short, scaling (or scale) is not one thing. By taking an agnostic position, we have been freed to investigate what and how scale is encountered in such projects.

I write this partially because it seems to me that there is often a hubris associated with mastery of a particular scalar device, as though one technique had somehow solved the problem of scale. Once, such pride was the prerogative of the quantitative statistician; today the tables have shifted and I hear such claims mostly from social network analysts – but watch out, big data is well placed to attempt a coup.

What I have attempted to show is that no single scalar device can serve to represent or manage all forms of scale. A network diagram of author collaborations will tell you a great deal about the composition of a domain, or the relation between disciplines. But for all the “completeness” of these data, it will tell you nothing about how many servers will be needed to manage a website’s traffic. And in turn, traffic indicators will not tell you why the communities of hydrologists and environmental engineers may chafe at having to work under the same umbrella; something a survey did moderately well.

Each of these approaches to scale provides a unique view and suggests different kinds of solutions to manage oncoming problems. But each device also generates exclusions and invisibilities.

In concluding my discussion of WATERS above, I mentioned Laura and Bob, real hydrologists with pseudonymous names. Laura trudges into wetland fieldsites to collect and document her data on paper. She was irked that there was no place for her to recount this in a survey that assumed all data were digital. Bob uses surrogates, such as a declined average of eggs per nest, as a signal of pollutants in a river. There was no space for him to describe this rich method in the ‘other’ text-field available in the survey. A survey is a powerful device for capturing collective similarities and differences, but even the most well-crafted survey apprehends only particular kinds of relations.

Laura and Bob know they are outliers – largely invisible to that scalar device – but they still wanted to have a voice in the design of the future infrastructure for their field. Their stories were captured instead by this ethnographer, recounted here in this paper, and in a handful of presentations back to WATERS PIs.

Ethnography too is a scalar device in such enterprises. This is why infrastructure developers invite investigators such as myself on-board. They hope that we will be able to help them understand things like the epistemic cultures of their user communities, to mitigate the difficulties of interdisciplinary collaboration, or to shape robust routines of communication.

Am I arguing that ethnography is somehow a better or less reductive scalar device? I am not. As with any other device, ethnography renders a particular representation of scale, myopic and revealing both.

In my experience, infrastructure developers implicitly agree. No single scalar device will suffice. They are forever seeking additional advice on their communities, better ways of organizing meetings, new forms of data visualization, and more agile methods for summarizing and communicating heterogeneous work. No approach is treated as a magic bullet, as if one universal representation of the scale of the enterprise will become available to topple all the others. There is no one scalar device to rule them all.

I have sat in rooms with principal investigators (or, information managers), as they poured over stacks of indicators in preparation for a decision. Multiple scalar devices are arranged idiosyncratically on tables and screens in order to stitch together representations of their enterprise. From a formal standpoint, the methods that generated such indicators may be wholly incommensurable: notes from an All-Hands Meeting, user requirements scenarios, results of a survey, Gantt charts, and a handful of accumulated statistics. But in such rooms – what Janet Vertesi has identified as “seamful spaces” [42], in which local and practical solutions to the challenges of integration are enacted – no formal methodological justification is needed to mix and match these indicators. It is at the intersection of multiple scalar devices that decisions are made.

**CONCLUSION:** Ethnography of scaling... “Yes, but does it scale?”

This paper opened with the most common methodological question I receive as an ethnographer studying
sociotechnical systems: “how can ethnography study large things?” I now conclude with the second most common question I get when I recount my ethnographic method, one that is more dismissive and exhibits less curiosity: “Yes … but does it scale?”

I have always been baffled by this question. When that question is asked, what come to mind are all the indicators produced about infrastructures that I have collected: troves of surveys and questionnaires, social network analyses, and benchmark metrics. I recall all the meetings where findings were presented, and the debates and conflicts that ensued; many of these events are documented only in my field notes. I also consider all the ways I have tracked these indicators as they were deployed to inform design decisions about a future sociotechnical system. Does ethnography scale? My answer is that I endeavour to scale exactly to the extent as do my informants.

At the root of this question, it seems to me, is a confusion about the object of ethnographic investigations. Scaling activities in all their forms are a crucial feature of infrastructure development, and sociotechnical systems more generally. One cannot observe activities of development without continuously encountering such indicators.

Perhaps because many scalar devices are quantitative or technical, there is an assumption that the (qualitative) ethnographer will skirt such activities. This is not the case. In the study of an enterprise where actors seek to represent and manage its size, scalar devices are a part of the field of observation for the ethnographer. Put more strongly, they should be objects of investigation.

As ethnographers our own methods should match our actors’ nimbleness. In treating organizational action and technical design as practice, talk and material arrangement, scaling is rendered observable to the analyst. There is nothing in advance to limit ethnographic studies to those questions that are deemed of the “correct size” or “sufficiently social.” I hope that the approach described here will open a terrain of research in studies of sociotechnical systems. We should travel across the scales mirroring the footsteps of our informants’, and when we do so we will discover that scaling is the sometimes mundane, but often esoteric, work of actors. Rendered observable as a form of practice, any size-complex of the qualitative researcher dissolves.

One final note on scalar agnosticism. My first reaction when encountering indicators of scale is to treat them in the fashion described in this paper: as an activity of scaling oriented to knowing and managing the size of the object. However, once my investigation of these activities of scaling is completed, I remove the “bracket” that is my scalar agnosticism, and stuff these indicators into a black box, just as the actors do. Thereafter, as modalities are stripped from scalar devices [26], I treat visualizations and summaries in the way that developers do: largely as facts. In addition to the practices of scaling, there are many other activities that are relevant to understanding the development of infrastructure. And infrastructure is large. The statistics and visualizations they generate help me to know and manage my investigations of those infrastructures.

ACKNOWLEDGEMENTS

Earlier versions of this paper were presented at ‘Scientific Practice as Ordinary Action’ in Fribourg, SW; ‘Mind and Society’ at the University of Manchester, UK; ‘The Ethnographer in the Network’ at Princeton University, US, and an informal workshop at the Norwegian University of Science and Technology. I would like to thank the participants at those events for their feedback, to Marte Giskeødegård and Jennifer Feldman for their insightful comments on final revisions, and to Anat Leibler for comments on very early drafts. This research has been supported by NSF/VOSS award #0904145 ‘Delegating Organizational Work to Virtual Organization Technologies.’

REFERENCES