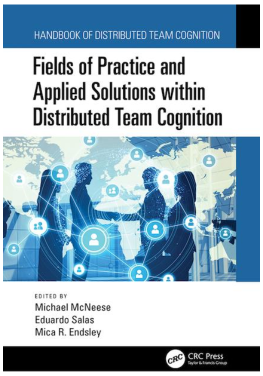

Fields of Practice and Applied Solutions within Distributed Team Cognition

Michael D. McNeese, Eduardo Salas, Mica R. Endsley

Stabilizing Digital Infrastructures in Distributed Social Science Collaboration

Publication details

https://test.routledgehandbooks.com/doi/10.1201/9780429459542-2

Steve Sawyer, Jaime Snyder, Matt Willis, Sarika Sharma, Carsten Østerlund, Emma Allen

How to cite :- Steve Sawyer, Jaime Snyder, Matt Willis, Sarika Sharma, Carsten Østerlund, Emma Allen. 29 Sep 2020, *Stabilizing Digital Infrastructures in Distributed Social Science Collaboration from:* Fields of Practice and Applied Solutions within Distributed Team Cognition CRC Press
Accessed on: 28 Nov 2023

2 Stabilizing Digital Infrastructures in Distributed Social Science Collaboration

Steve Sawyer, Jaime Snyder, Matt Willis, Sarika Sharma, Carsten Østerlund, and Emma Allen

CONTENTS

We focus on the uses of digital resources by social scientists collaborating at a distance in order to pursue insights into "infrastructure in use." In doing so we advance the concept of stabilization as a sociotechnical activity that demands extensive negotiations relative to the structure, form, uses, and arrangements of digital materials. And, we situate the work of stabilization relative to the extensive literature and developing conceptual basis of distributed cognition (Hutchins, 1995; Salas & Fiore, 2004; Fiore & Wiltshire, 2016; Nemeth, O'Connor, Klock, & Cook, 2006; Stanton, 2016).

Four reasons motivate us to pursue this work. First, we seek to articulate the sociotechnical concept of *stabilization* with respect to the more technocentric concept of *standardization* as an organizing principle for assembling the digital resources supporting distributed collaboration. We see stabilization as a negotiated order that brings together human and non-human participants into a more coherent arrangement. We contrast this to standardization or the reliance on technical standards to encourage interoperability among particular technological elements, with the people tethered (willingly or not) to such arrangements.

Second, as part of distinguishing stabilization from standardization, we articulate stabilization as being an example of distributed cognition, and seek to connect these literatures (e.g., Nemeth et al., 2006; Stanton, 2016). Scholars of human factors and particularly those focused on the complexities of distributed cognition involving material and digital arrangements have developed a rich and robust conceptual framework that bears directly on stabilization, distributed scientific work, and current thinking on cyberinfrastructure (CI) (Hutchins, 1995; Salas & Fiore, 2004; Fiore & Wiltshire, 2016; Nemeth et al., 2006; Stanton, 2016).

Our third reason for advancing stabilization practices as core to distributed work is that contemporary scientific endeavors are increasingly differing from traditional approaches to science in that projects are growing in terms of both budget and effort, resulting in larger and more distributed teams and a greater reliance on computing and other digital arrangements, typically referred to as CI (Atkins et al., 2003; Foster & Kesselman, 2003). There is now a visible body of testimonial literature about this trend (Atkins et al., 2003; Hey & Trefethen, 2003; Katz & Martin, 1997); a growing collection of empirical literature providing insights into movements toward CI-enabled science practice (cf. Ribes & Lee, 2010); a nascent area of literature providing conceptual insight into these practices (Jackson, Edwards, Bowker, & Knobel, 2007; Ribes & Finholt, 2009); and a history of various design approaches to designing and building the digital platforms supporting computer-enabled science (e.g., Borgman, Bowker, Finholt, Arbor, & Wallis, 2009; Jirotka, Lee, & Olson, 2013; Lee, Dourish, & Mark, 2006).

Our fourth reason for pursuing this work is that the burgeoning literature on CI seems to be predicated on an implicit model of science where roles are well-defined, scientific workflow is stable, discrete, and modular, and the value of these digital infrastructures lies in their provision of high-value assets (such as complex, rare, or sizable combinations of multiple data sets, analytic devices, and other massive-scale resources) to a large, distributed, and often diffuse number of scientists (Atkins et al., 2003). This model of scientific practice reflects a socially thin, compartmentalized, and streamlined version of scientific work. In contrast, research and reports on the observed collaboration practices of these scientists make clear that distributed work in this domain also depends on building long-term relationships of trust and mutual understanding (e.g., J. Cummings, Finholt, Foster, Kesselman, & Lawrence, 2008; Star & Ruhleder, 1996; Haythornthwaite, 2009). The empirically centered work presented here helps make clear that scientists using CI often rely on significant social supports and the sustained engagement of colleagues. That is, there is substantial social activity to move the scientific endeavor forward, what Haythornthwaite (2009) articulates as "heavyweight" social activity, requiring extensive negotiation and discussion, and reliant on building and sustaining shared understandings.

In this chapter we build from our studies of the work practices and uses of digital resources by social scientists collaborating at a distance in order to advance empirical and theoretical insights into the ways in which infrastructural stability—the efforts to ensure that digital resources and work practices stay aligned—support science practices. Specifically, we make the case that the sociotechnical concept of stabilization provides a more robust view of how scientists use CI than does the more well-known, and technology-centered, concept of standardization. Stabilization reflects the hard work of negotiation, adaptation, and shifting work practices that allow those using a particular technological system (in this case scientists using CI), to settle on a particular arrangement (see Russell & Williams, 2002). Stabilization embodies a balancing of competing technological and social forces—different possibilities of use, different goals of the users and designers, and emerging possibilities of use. In contrast, the concept of standardization focuses attention to the ways in which specific technological arrangements are designed to allow the elements to interoperate on a systems level. Standardization is what allows multiple devices to be connected together, and for documents to be visible and editable on different digital platforms. Standardization is a tremendous engineering feat, even as these often diminish the ways in which people organize and leverage the standards.

Building from this, we set out here to theorize on the role of stabilization in distributed work, and the ways in which these efforts are enmeshed into the digital and material technologies used by distributed teams. In the next section we lay out the literature relevant to studying social scientists' shared uses of CI. Then we lay out our research design, data collection, and data analysis. In the third section of this chapter, we highlight the forms of CI we observe scientists using. In the final section we explore implications of stabilizing for both research on scientific collaboration and for CI.

STUDYING USES OF CI BY SOCIAL SCIENTISTS

Social science enlists those whose central phenomenon of interest is the human condition, many of whom are found in disciplines such as: sociology, economics, anthropology, political science, linguistics, management and organizational studies, and others (Goff et al., 2011). Motivated by a gap in current research on the uses of CI to support science, we sought to explore the practices of social scientists involved in distributed collaborations. This exploration of social scientists' uses of digital resources enables us to compare the work of social scientists with what has been found in studies of natural, physical, and computer sciences (cited previously and detailed in the following sections), while also identifying variations among social scientists based on training, research methodologies, and personal preferences. And, the standardization of infrastructure within these CI environments is greatly influenced by the research practices, experiences, and uses of digital tools and material resources, and the data structures associated with diverse teams of scientists collecting and working with large data sets over extended periods of time (Ribes & Lee, 2010; Baker, Ribes, Millerand, & Bowker, 2005)

To do this we review three distinct areas of research. We begin by focusing on CI, acknowledging the implicit orientation towards CI's for the natural and physical sciences. Next, we describe the challenges of CI for distributed teams from a work practice perspective. In the third section we describe the enacted view of CI that focuses attention to the ways in which scientists currently learn and understand CI. Finally, we highlight what a practice-based perspective on distributed collaboration can contribute.

CI AND SCIENTIFIC WORK

Contemporary CIs are typically motivated by a need for some combination of (1) computational power, (2) data storage and access, and (3) shared uses of important (and often expensive) data collection, data analysis, and visualization platforms. Because of these needs, CIs often provide high-speed networking technologies to support computation, data access, and tool access. These CIs often are premised on large teams (i.e., dozens, if not hundreds, of scientists) with clearly defined roles, conducting distributed, collaborative research on pressing scientific problems. Increasingly, CI-enabled science is becoming the norm in the natural, physical, and biological sciences. Standardization of infrastructure within these environments is greatly influenced by the scientific practices and standardized data structures associated with diverse teams of scientists collecting and working with large data sets over extended periods of time (Ribes & Lee, 2010).

Current CIs in astronomy, bioinformatics, geosciences, ecology, medicine, and environmental studies have at least three things in common.[1] First, CI systems are typically designed for streaming data or the colocation of data to study a large-scale phenomenon. Second, CI-enabled work is long term and wide reaching with multiple stakeholders (Steinhardt & Jackson, 2015). Third, the goal of a CI is to support not just digital resources like data but also physical resources along with computationally intense environments for collaborative research work (Ribes, 2014). Fourth, CIs are supported by human infrastructures that manage the organizational and computational work needed to sustain them over time (Lee et al., 2006; Bietz, Baumer, & Lee, 2010).

Empirical insights from studies of CI uses make clear that much of the value of CI in these contexts is centered on gaining access to data. Baker, Ribes et al. (2005) argue that achieving data interoperability is the way to bridge intellectual communities. Likewise, sharing data on a large scale is seen as the key way to bring together interdisciplinary science to solve problems; this is intuitive but not well documented or systematically investigated (Borgman et al., 2009). Moreover, it appears that scientists are shifting their practices to leverage these resources. For example CI-enabled collaborations are increasingly being designed for leveraging large data sets, large-scale computing resources, and high-performance visualization tools, using collaborative workflows and allowing several people to share and work on the same analyses while located at different places (Jirotka et al., 2013).

WORK PRACTICES IN DISTRIBUTED SCIENTIFIC WORK

Distributed collaboration is increasingly common in scientific work (Atkins et al., 2003; Hey & Trefethen, 2005; Karasti, Baker, & Millerand, 2010) and lies at the heart of CI efforts (Borgman et al., 2009). Distributed collaborative scientific work refers to "interaction(s) taking place within a social context among two or more scientists that facilitates the sharing of meaning and completion of tasks with respect to mutually shared, subordinate goals" (Söderholm et al., 2008).

These efforts rely on complex social relationships that are often tightly interwoven into the scientific and technological practices of collaborators (e.g., Lee et al., 2006; Haythornthwaite, 2009). Scientists may choose to collaborate because they do not have sufficient skills or resources to pursue the work independently, they see value in drawing in other perspectives and talent, they are incentivized through funding and other reward structures to partner, or for other reasons

(Katz & Martin, 1997). Collaboration, more broadly, is fundamental to science as it is the basis of peer review and the epistemic cultures on which they rely (e.g. Knorr-Cetina, 1999, 2009)

Why does distributed scientific collaboration require complex social relationships? Beyond the technical difficulties of achieving data interoperability, Ribes and Lee (2010) identified four current challenges for developing CI to support distributed scientific collaboration: (1) heterogeneity of work practices and social structures among collaborators, (2) the need for standardization of data structures, work practices, and analytic approaches, (3) the importance of designing to allow changes over time, and (4) the need to design CI to be sustainable (and maintainable) over time scales of decades, what Ribes and Finholt (2009) call the "long now" of technological infrastructure (see also Karasti et al., 2010). Each alone is quite challenging; that all four of these, and the efforts to ensure data interoperability (e.g. Baker & Yarmey, 2009), are happening together magnifies the complexities involved in maintaining these collaborative practices.

Supporting scientists who work at a distance requires ICTs that make it possible to interact routinely, to easily access data repositories, and to support opportunities for connections with other scholars that overcome time and space (Olson & Olson, 2000; Crowston, Specht, Hoover, Chudoba, & Watson-Manheim, 2015). The evidence makes clear that coordinating these activities requires heavyweight social effort and comes at a substantial cost in both time and money (J. N. Cummings & Kiesler, 2008). This social effort is nonlinear, as a function of the number of dyads among all the project members, the level of turnover of members, and—to some extent—the complexity of the shared work. And, given both the changing nature of scientific enterprises and increasing levels of collaborative scientific effort, CI scholars emphasize the concept of the "long now" or the importance of designing for an open future (Ribes & Finholt, 2009). Changes over time in project institutional affiliation, membership, and work practices need to be accommodated in order for a CI to be sustainable across scientific inquiry (Bietz, Ferro, & Lee, 2012).

Jackson et al. (Jackson, Ribes, Buyuktur, & Bowker, 2011), building on a multi-year, multi-site, ethnographic field study of CI uses, found several issues related to time were also problematic in distributed collaborative scientific practice (see also Watson-Manheim, Chudoba, & Crowston, 2012). Jackson et al. observed that scientists had difficulty working with collaborators from institutions who have different academic calendars. More broadly, temporal rhythms of one's local work practices, organizational calendars, and scientific deadlines (for grants, conferences, and meetings) were critical shapers of distributed collaborative practices (e.g., Orlikowski & Yates, 2002).

A CI's sustainability rests on an ability to maintain standards, therefore one solution for addressing these challenges is standardization of data, work practices, and digital infrastructures (Ribes & Lee, 2010). Studies have highlighted the importance of data standardization and interoperability (Edwards, Mayernik, Batcheller, Bowker, & Borgman, 2011; Hey & Trefethen, 2005), while others have articulated the importance of, and difficulties with, standardizing scientific and analytic practices (e.g., Hara, Solomon, Kim, & Sonnenwald, 2003; Ribes & Lee, 2010). Still, the importance of supporting heavyweight social activities in distributed collaborations and addressing challenges related to standardization of CI in this domain is driving scholars to more carefully examine the evolving interdependencies among scientists, work practices, digital and material resources, and infrastructure.

CI PRACTICES

We conceive of scientific work as emerging from enacted practices (i.e., the ways in which people *do* science in specific contexts) that are embedded within organizational and material structures including digital and material resources (see also Star & Ruhleder, 1994). Seen this way, infrastructure like CIs *are* practices: bound up in doing, not simply being present. That is, a CI is fundamentally a sociotechnical arrangement: a combination of material and digital resources entwined with the activity of doing science (Ribes & Lee, 2010). This practice-based perspective focuses attention on the *doing of infrastructure* in science work. As such, digital infrastructures are organic entities that emerge out of “information and work needs” (Feldman & Orlikowski, 2011; Østerlund & Carlile, 2005; Star, 2010).

Seen this way, CI are enacted, their digital and material properties become involved in science as the collective product of shared forms of practices and help to demarcate the specific ways scientists engage with one another. In related work, Sawyer, Crowston, and Wigand (2014) explore the computerization of real estate agents’ work practices, focusing specifically on the ways real estate agents drew on various elements of existing digital infrastructures in ways that made sense to them. Sawyer et al. characterize the ways in which realtors perform digitally enabled work by

> drawing on many different computing elements and software based systems selected by individuals that may not be well-integrated or formally planned . . . [and] is thus ad hoc and only governed to the extent that the individual doing the assembling is making choices.
>
> **(p. 51)**

The resulting “infrastructure in use,” while not reaching the level of formalization provided by a CI, is functionally equivalent to a standardized information system, but retains flexibility (Ribes & Polk, 2014) to respond to a community member’s evolving work practices and material resources (Pollock & Williams, 2010; see also McNeese, Rentsch, Burnett, Pape, & Menard, 1998). As shown by the scientists in Vertesi’s (2014) study and Sawyer et al.’s (2014) real estate agents the doing of infrastructure relies on learning practices from peers. This has less to do with learning formalized protocols or procedures and more to do with discovering ways to work *with* and *through* stabilized digital assemblages.

RESEARCH DESIGN, DATA COLLECTION, AND ANALYSIS

Because of our interests in advancing current empirical and conceptual understanding of social scientists’ work practices and uses of CI resources, we designed an exploratory study (Punch, 2013). Our goal was theory elaboration: to advance the depth and clarity of current conceptualizations of cyberinfrastructure (e.g., Dutton, 2007; Ribes & Lee, 2010). Studies of scientific CI in use highlight that these digital resources are important and that standardization of data structures and research practices is one of the reasons for this. To advance this theorizing we designed the data collection in ways that engaged current scientists, that focused on work practices and the use of digital resources, and that focused on typically sized collaborative efforts (see also Sawyer, Østerlund, & Kaziunas, 2012).

DATA COLLECTION

Data collection methods consisted of a combination of semi-structured interviews and a detailed survey focused on digital tools and practices. The four-part semi-structured interview protocol was designed to be conducted face-to-face or via telephone. We asked respondents to (1) describe current distributed collaborations; (2) provide specific details about one of those projects; (3) describe current work practices on that project; and (4) detail the particulars of the digital elements (hardware, software, and services) used to support distributed collaborations. For each section, there were three to five broad questions with a number of probes.

We developed a survey to supplement the interview by capturing details of specific tools and patterns of use that might have otherwise been lost in narratives recounting specific projects, collaborative relationships, and challenges faced in conducting distributed social science work. The survey included a comprehensive "check sheet"—a form in which we asked participants to share information about collaborators as well as indicate in some detail specific digital devices, software tools, and other infrastructural elements that they used. The interview protocol and survey were designed so that the interviewee could work on the check sheet while we worked through the interview questions. The survey was available both on paper and online, with delivery mode dependent on the resources available at the time of interview and preferences of respondents.

Relying on several rounds of snowball sampling, we conducted interviews with 31 social scientists across one calendar year. Interviews took 70 minutes, on average. Interviewing took place in two extended phases (of 14 and then 8 interviews) separated by several months of interim analysis. During the interview, each social scientist shared details about a single distributed collaboration in which he or she was involved. Reports of collaborative practices provided to us by our primary informants gave us insights into the work practices of over 170 researchers and project staff (with all but two working on teams of between four and seven people). The 31 interviews represent the experiences of 8 men and 14 women pursuing social science questions in the intellectual communities of science and technology studies, information science, information systems, and HCI/CSCW. Nine of the 31 were tenure-track faculty and 14 were tenured faculty at research-intensive institutions in the United States and abroad; eight were research faculty or post-doctoral researchers. We note this distribution of experience (and likely experience) in our discussion of the findings in the next section.

DATA ANALYSIS

We recorded and transcribed the interviews, made digital copies of the check sheet to support analyses, and gathered particular documents and other project material (such as websites) when allowed. Transcripts of the 31 interviews were analyzed inductively to identify common themes related to particular aspects of (1) work practices of collaborative teams and (2) the ways in which these teams interacted with their CI and other digital technologies. To pursue the first focus, we began with a nominal list of practices and modified the initial list based on the data. To pursue the second focus, we coded the details of particular technologies being used as a means of identifying patterns of technological arrangements that were common to multiple teams. We also focused on the ways in which participants articulated uses of CI and technologies as they discussed practices and issues.

Two members of the research team coded each interviews and then worked through several iterations of a standardized coding scheme with the entire research team to clarify emergent themes, resolve differences, better articulate formative codes, and add new concepts to the analytic scheme. Typical of open coding on interviews and field notes, during this data analysis, consensus was valued more highly than inter-rater reliability (Glaser & Strauss, 1967).

TWO FINDINGS: NASCENT CI AND STABILIZING PRACTICES

We focus on stabilization practices, a topic that emerged from the data (it was not one of the initial practices being coded). Stabilization took the form of efforts to adapt to changes in the features and functions of specific software, issues with compatibility across different software applications and suites, changes in research collaboration member's collective or individual expertise with software (due to changes in personnel or changes in software), shifts in team member's understanding of the working arrangements, learning effects when decisions were made to switch software (for a variety of reasons), and how people recovered from mistakes with version control and document management.

To better understand these stabilization efforts, we next report on patterns of digital technology uses to make clear four insights from the data. First, data show there is no standard arrangement, no singular set of digital technologies, being used by respondents. Second, data make clear that there are many combinations and overlapping uses of similar software and applications, that these arrangements evolve for a number of reasons, and that this change and overlap is expected by the participants. Third, data make clear that interoperability—enabled by standards—among the software, applications, and platforms relies on many technical standards (such as CSV and XML for sharing data, PDF and other standard document formats for sharing files, and many standards for email, web, and internet working). Fourth, we observe that these collections of digital technologies, while having similar functionality, being comprised of multiple commercial and open-source applications and platforms, and all being used to support distributed scientific collaboration, fall short of being an enduring CI. As such, standardization is a necessary technological requirement for these proto-CI arrangements, but is not sufficient to ensure these digital arrangements are workable or useful for the collaborating scientists.

As readers might expect, participants reported having and using several computers (both laptops and desktops), tablets, and smartphones in support of their work. They used these various devices at home, at various places at work, in public spaces, and in between. Several had print servers at home. Most digital devices were provided by an employer, although each respondent noted they also owned several personal devices. And, respondents indicated that they routinely used both personal and employer-provided devices together. Perhaps unsurprisingly, email was the dominant means for interacting, collaborating, sharing materials, and for almost all project work. While many of our respondents were active in social media, they reported that email was still the common denominator, omnipresent in their lives and accessed across multiple devices: “A lot of people still email things around even though there’s a Dropbox” (JMp6).

These digital resources were used every day, often multiple times. In describing their uses of devices, most respondents found it challenging to separate activities directly related to the specific collaborative project being described, similar activities being performed for different projects, and personal use. This noted, only some of these uses pertain directly to their distributed project collaborations. However, emails pertaining to student grading, service assignments, and data analysis arrive to the same device. What this means is that scientific practices and access to digital scientific resources were embedded into the larger sphere of their daily life and both science and non-science happened across the same set of digital and material resources, mixed together across time. None of the 31 respondents reported having distinct labs as the site of their research. While most had one or two preferred places to work (their office at home or a favorite cafe), in practice their research happens at different times in their personal schedules and across many physical places, even in public spaces like airports as they travel. Collectively, through doing these practices, scientists enact and recreate the shared distributed cognition that defines how they work.

Respondents identified using nearly 70 distinct software applications and platforms. These included common commercial software and hardware such as Microsoft (MS) Office applications, the Google suite of software, standard browsers such as Firefox or Safari (but rarely Internet Explorer), typically running on Microsoft or Apple computers. We note that all of these are either commercial products or open-sourced; none are purpose-built CI components. The teams and collaborators who provided us data relied on commercial and community offerings, a “commodified” approach to technology development that relies on standards and, for the commercial products, market competition (Sawyer, 2000).

Most of these commercial products provide a means to share data and documents across rival platforms. This means that team members can choose to use different word processing software and exchange files, or to experiment with particular cloud-storage providers, or move from SPSS to *R*, knowing that these are more or less compatible, more commodities than specialized.

Nineteen respondents report using Gmail; 21 also use a second email address and/or client (such as Thunderbird, Outlook, or Apple Mail). Most are using no more than two email clients, though there are 67 different email uses across the 31 respondents. All but one respondent use MS Word, and 22 use Google Docs, which means that most respondents use both. Twenty-seven respondents reported using some form of reference management software (e.g., Endnote (12), Zotero (8), Mendeley, and others), with several people using more than one. This noted, no one reference management software application had a majority of use. Likewise, there was a wide range of analytic software being used (with R having nine counts; SPSS, SAS, Atlas TI, and Dedoose all having four counts; and many with one count). If a respondent reported using analysis software, they reported more than one. Only three of 31 interviewees mentioned doing collaborative data analysis (on numerical data); the rest participated in collaborations where each individual did their own analysis and shared or pooled their results.

All respondents reported using one or more cloud-based repository such as Dropbox, Google Drive, or Box to store and share project documents. No respondent relied on his or her institution's learning management system or on private infrastructures (like personal web space) to serve this function.

Skype served primarily as a vehicle for project meetings during which updates, to-dos, and plans were discussed, though ten people also noted they used call-in conference phone services. Participants reported using a range of project management software (but only by 13 of the respondents). Most had one use. Some reported using note-taking software (e.g., Evernote). Five participants reported using GitHub (in the project management category).

Participants spoke at length of their efforts to make the various elements of the proto-CI in use work. Some of the issues reflect the realities of having to know many pieces of software. For example, having to discuss how to move numerical data from Excel to R, or how best to format a Google-Doc-based document so that it can be copied into a Word document. This form of articulation work is common outside of scientific collaboration and independent of a reliance on CI. Stabilization differs from articulation in that it requires a collective negotiation: multiple participants are involved and the outcome or decision requires the team to adhere to the new arrangements (whereas articulation work is simply extra work, often done by one person outside the view of others).

Stabilization practices show up in both the mundane activities of working together, often most visible when personnel change and issues arise. Two constants of these collaborative efforts, reported on by all 31 respondents, are (1) the steady flow of students (both graduate and undergraduate) on and off of the research teams and (2) collective attention to reflecting on how the collaborators are following protocols and using specific digital technologies. The steady changes in team composition lead to constant discussions about basic procedures (how do we name files, where are they placed, who takes notes, who manages the to-do list) and larger tasks and goals and a re-establishment and redistribution of shared cognition (Hutchins, 1995; Salas & Fiore, 2004; Fiore & Wiltshire, 2016; Stanton, 2016). Relative to the basic procedures, the proto-CI elements are involved in subtle and not-so-subtle ways: updating which citation management system to use based on losing a high-skilled student for one who has limited skills on one application; getting students to use Dropbox instead of Google Docs; or convincing new graduate students to share notes in a common folder.

Relative to stabilization efforts around larger tasks and goals, the entire team gets involved when a senior doctoral student moves on, as they collectively must re-sort who does what and which digital platforms to use. Stabilization efforts often frame decisions on projects, tasks, and to-dos; it is an omnipresent effort.

STABILIZING NOT STANDARDIZING: IMPLICATIONS FOR RESEARCH AND PRACTICE

Through this research we come to recognize that customized assemblages of standardized tools create a *stabilized* (but not static) infrastructure that enables collaborative teams to maintain flexibility in their work practices throughout the evolution of their research project. Second, this shift of focus from standardization to stabilization in the context of the heavyweight collaborative activities that lead to distributed cognition has implications for how we think about infrastructure design to support scientific work. Designs for CI need to enable scientists to build and maintain shared mental models (e.g., Stanton, 2016).

These observations echo findings from other empirical studies of CI use and distributed scientific collaboration (Fry, 2006; Lin, Procter, Halfpenny, Voss, & Baird, 2007; e.g., Ribes & Lee, 2010). The social scientists in our study described their efforts to *assemble*, rather than standardize, their digital infrastructures to accommodate their heterogeneous—and often idiosyncratic—preferred work and social practices. Teams relied on commodified software and commercial digital networks and storage providers, working digital infrastructures, relatively short time horizons, and stronger commitments to relationships and practices over specific hardware or software. These collections of commodity technologies point to stable, but ad hoc, sets of arrangements, more reflective of the digital assemblages of real estate agents discussed previously (Sawyer et al., 2014) than of CI. These bundles of ICT are not an information system in the classic sense: there is no governance, no defined coherence; rather, functional equivalence driven more by the similarity across these respondents' work roles and practices than by any a priori design (Sawyer et al., 2014).

Given these observations, we argue that the practice of managing the confluence of infrastructures and collaborative practices that we see in the work of distributed social scientists is more closely associated with technological *stabilization* (Pinch & Bijker, 1987) than *standardization*. In discussing mechanisms of technology development, Pinch and Bijker challenge the idea that technologies develop as a result of a predetermined march towards standardization. In Pinch and Bijker's (1987) conceptualization of stabilization, new technologies undergo a period of interpretive flexibility marked by multiple options and incarnations followed by a phase of social integration that ultimately results in a stabilizing convergence around a particular arrangement. They provide the example of the development of the bicycle. Some early bicycle designs failed and others thrived. Success or failure was not based on a linear evolutionary model where one design arose from the previous, but rather as a result of multidirectional experiments that enabled makers to explore a range of different options within the two-wheeled peddled vehicle design space, each responding to distinct problems faced by different user groups. Eventually designs stabilized into the safety bicycle (two wheels of equal size, air tires and a seat in a more or less forward position), a form that has remained relatively constant ever since.

We recognize in this description the assemblages of the social scientists in our study; different components of a digital infrastructure are gathered in response to specific problems being faced by a team at a given time in their research and collaborative process. The beginning phases of a project reflect the interpretive flexibility described by Pinch and Bijker (1987) during which different tools and practices are evaluated and considered. This is followed by a stabilizing convergence during which a team settles on a given configuration of tools. However, where Pinch and Bijker (1987) talk about the process of technological stabilization as being completed once and done, we saw evidence of the teams in our study returning to this process repeatedly during the course of a project's lifetime, re-interpreting or re-negotiating the specifics of their infrastructure assemblage in response to changes in research practices, personnel, and/or resources.

In this sense, the stabilizing practices of the social scientists in our study evoke Vertesi's (2014) "seamful spaces." She draws a compelling contrast between the goal of "seamless integration" envisioned by ubiquitous computing researchers and the reality of "seamful" digital environments in which many distributed scientific collaborations exist. In practice, social scientists are continually assessing and responding to changing infrastructural needs at technoscientific, sociotechnical, and institutional levels, three dimensions of change described by Ribes and Polk in their discussion about infrastructure flexibility (2014). We examine this process in more detail next.

Ribes and Lee describe the ways in which CI supports science "by identifying vast interdisciplinary swaths that could benefit from data and resource sharing, knowledge transfers, and support for collaboration across geographical, but also institutional and organizational divides" (Ribes & Lee, 2010, p. 232). We previously highlighted some of the challenges that CI designers face when trying to accomplish this, including: (1) the heterogeneity of work practices and social structures among collaborators, (2) the need for standardization of data structures, work practices, and analytic approaches, (3) the importance of designing to allow changes over time, and (4) the need to design CI to be sustainable (and maintainable) over long time periods (e.g., multiple decades).

Based on what is reported in current CI research, we anticipated that our respondents would use standardized tools in consistent ways across projects because they provided the necessary digital infrastructure to support their social science practices. However, as we learned more about the work practices of these distributed collaborating scientists, we discovered that standardized tools and infrastructures, such as email, Google Docs, and Dropbox, were often used in different ways by each team of researchers. The science being conducted by our participants' teams required sustained engagement, flexibility, and responsiveness to fellow team members. We identified this as a culture of accommodation, similar to what Schroeder and Spencer (2009) have previously reported: social scientists are both inventive and adaptable. We observed a high degree of interdependence among the processes for defining roles, negotiating rules or routines, and articulating tasks. Many of the stabilizing and adaptive actions described were motivated by a need to support valued heavyweight aspects of teamwork better conceptualized as distributed cognition

(Hutchins, 1995; Salas & Fiore, 2004; Stanton, 2016). Overall, the respondents' individual science practices were loosely and often uniquely mapped to the shared practices of a specific distributed scientific collaboration (see also McNeese et al., 1998).

We also note that others have found that sustaining data access and interoperability is difficult for CI (e.g., Baker, Jackson, & Wanetick, 2005; Bietz et al., 2012; Lawrence, 2006; Lee et al., 2006; Barjak et al., 2009); therefore, we highlight the opportunity to develop tools that will help transition some research tasks to lighter-weight environments through responsive workflow support (e.g., Fry, 2006). Further, training regarding how to collaborate is typically learned in apprenticeship, by watching faculty move from project to project and mimicking this craft as part of scholarly maturation. In order to create more intentionally around these adaptive and stabilizing practices, it will become increasingly important to train PhD students on the fundamentals of distributed collaboration, to help them learn community or shared data practices (e.g., 2007) and to develop basic technical skills that will enable them to practice adaptive design to modify and optimize their collaborative environments. Part of this training will need to involve cultivating the ability to assess, implement, and evaluate the effectiveness of digital tools in terms of functional equivalence.

ACKNOWLEDGMENTS

Funding provided by National Science Foundation (NSF) via Grant CI 1527410. The findings and opinions reported are solely the authors and do not represent or speak for the NSF.

We gratefully acknowledge the time and insights provided by so many scholars, and the comments from reviewers and audience members of the various talks, panels, and feedback sessions we have held. Thank you to Michael D. McNeese and Eduardo Salas for specific comments on earlier versions of this chapter.

We thank the many students who have contributed so much to this program of study over the years, to include: post-doctoral fellow JoAnn Brooks; doctoral student Sarah Bratt; master's students Elizabeth Kaziunas, Doratea Szkolar, Michelle Brown, Megan Threats, Elizabeth Greenberg, and Nirihika Ved; undergraduate student Nicholas Kapteyn; and post-baccalaureate research fellow Anjelica Torcivia.

NOTE

1. This is an illustrative list; there are many forms of CI to be found in many intellectual spaces not mentioned.

REFERENCES

Atkins, D. E., Droegemeier, K. K., Feldman, S. I., Garcia-molina, H., Klein, M. L., Messerschmitt, D. G., . . . Wright, M. H. (2003). Revolutionizing science and engineering through cyberinfrastructure: Report of the national science foundation blue-ribbon advisory panel on cyberinfrastructure. *Science*, *81*(8), 1562–1567. Retrieved from www.nsf.gov/od/oci/reports/atkins.pdf

Baker, K. S., Jackson, S. J., & Wanetick, J. R. (2005). Strategies supporting heterogeneous data and interdisciplinary collaboration: Towards an ocean informatics environment. In *HICSS'05*. New York: IEEE Computer Society Press.

Baker, K. S., Ribes, D., Millerand, F., & Bowker, G. C. (2005). Interoperability strategies for scientific cyberinfrastructure: Research and practice. *Proceedings of ASIST*, *42*(1).

Baker, K. S., & Yarmey, L. (2009). Data stewardship: Environmental data curation and a web of repositories. *International Journal of Digital Curation*, *4*(2), 12–27.

Barjak, F., Lane, J., Kertcher, Z., Poschen, M., Procter, R., & Robinson, S. (2009). Case studies of e-infrastructure adoption. *Social Science Computer Review*, *27*(4), 583–600.

Bietz, M. J., Baumer, E. P., & Lee, C. P. (2010). Synergizing in cyberinfrastructure development. *Computer Supported Cooperative Work (CSCW)*, *19*(3–4), 245–242.

Bietz, M. J., Ferro, T., & Lee, C. P. (2012). Sustaining the development of cyberinfrastructure: An organization adapting to change. *CSCW'12*, 901–910.

Borgman, C. L., Bowker, G. C., Finholt, T. A., Arbor, A., & Wallis, J. C. (2009). Towards a virtual organization for data cyberinfrastructure. In *Proceedings of the 9th ACM/IEEE-CS joint conference on digital libraries* (pp. 353–356). New York: IEEE Computer Society Press.

Crowston, K., Specht, A., Hoover, C., Chudoba, K. M., & Watson-Manheim, M. B. (2015). Perceived discontinuities and continuities in transdisciplinary scientific working groups. *Science of the Total Environment*, *534*, 159–172 . https://doi.org/10.1016/j.scitotenv.2015.04.121

Cummings, J. N., Finholt, T., Foster, I., Kesselman, C., & Lawrence, K. A. (2008). *Beyond being there: A blueprint for advancing the design, development, and evaluation of virtual organizations.* Report from an NSF Workshop on Developing Virtual Organizations.

Cummings, J. N., & Kiesler, S. (2008). Who collaborates successfully ? Prior experience reduces collaboration barriers in distributed interdisciplinary research. *CSCW'08*, 437–446. https://doi.org/10.1145/1460563.1460633

Dutton, W. H. (2007). Reconfiguring access to information and expertise in the social sciences: The social shaping and implications of cyberinfrastructure. In *Proceedings 3rd international conference on e-social science, October 7–9*. Ann Arbor, MI: The University of Michigan. Retrieved from http://ess.si.umich.edu/papers/paper129.pdf

Edwards, P. N., Mayernik, M. S., Batcheller, A. L., Bowker, G. C., & Borgman, C. L. (2011). Science friction: Data, metadata, and collaboration. *Social Studies of Science*, *41*(5), 667–690.

Feldman, S. I., & Orlikowski, W. (2011). Theorizing practice and practicing theory. *Organization Science*, *22*(5), 1240–1253.

Fiore, S., & Wiltshire, T. (2016). Technology as teammate: Examining the role of external cognition in support of team cognitive processes. *Frontiers in Psychology*, *7*, 15–31. https://doi.org/10.3389/fpsyg.2016.01531

Foster, I., & Kesselman, C. (2003). *The grid 2: Blueprint for a new computing infrastructure*. Amsterdam, Netherlands: Elsevier.

Fry, J. (2006). *Coordination and control of research practices across scientific fields: Implications for a differentiated e-science*. Hershey, PA: Ed. Info. Sci. Pub.

Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research*. Hawthorne, NY: Aldine de Gruyter.

Goff, S. A., et al. (2011). The iPlant collaborative: Cyberinfrastructure for plant biology. *Frontiers in Plant Science*, *2*(34).

Hara, N., Solomon, P., Kim, S., & Sonnenwald, D. H. (2003). An emerging view of scientific collaboration: Scientists? perspectives on collaboration and factors that impact collaboration. *JASIST*, *54*(10), 952–965.

Haythornthwaite, C. (2009). Crowds and communities: Light and heavyweight models of peer production. In *HICSS'09* (pp. 1–10). New York: IEEE Computer Society Press.

Hey, T., & Trefethen, A. (2003). e-Science and its implications. *Philosophical Transactions of the Royal Society of London Series A Mathematical Physical and Engineering Sciences*, *361*(1809), 1809–1825. Retrieved from http://eprints.ecs.soton.ac.uk/7964/

Hey, T., & Trefethen, A. (2005). Cyberinfrastructure for e-Science. *Science*, *308*(5723), 817–821. Retrieved from http://eprints.soton.ac.uk/260844/

Hutchins, E. (1995). *Cognition in the wild*. Cambridge, MA: MIT Press.

Jackson, S. J., Edwards, P. N., Bowker, G. C., & Knobel, C. P. (2007). Understanding infrastructure: History, heuristics, and cyberinfrastructure policy. *First Monday*, *12*(6).

Jackson, S. J., Ribes, D., Buyuktur, A., & Bowker, G. C. (2011). Collaborative rhythm: Temporal dissonance and alignment in collaborative scientific work. In *CSCW'11* (pp. 245–254). New York: Association of Computing Machinery (ACM) Press.

Jirotka, M., Lee, C. P., & Olson, G. M. (2013). Supporting scientific collaboration: Methods, tools and concepts. *CSCW'13*, 1–49. https://doi.org/10.1007/s10606-012-9184-0

Karasti, H., Baker, K. S., & Millerand, F. (2010). Infrastructure time: Long-term matters in collaborative development. *Computer Supported Cooperative Work (CSCW)*, *19*(3/4), 377–415.

Katz, J. S., & Martin, B. R. (1997). What is research collaboration? *Research Policy*, *26*(1), 1–18.

Knorr-Cetina, K. (1999). *Epistemic cultures*. Cambridge, MA: Harvard University Press.

Knorr-Cetina, K. (2009). *Epistemic cultures: How the sciences make knowledge*. Cambridge, MA: Harvard University Press.

Lawrence, K. A. (2006). Walking the tightrope: The balancing acts of a large e-research project. *Journal of Computer Supported Cooperative Work*, *15*(4), 385–411.

Lee, C. P., Dourish, P., & Mark, G. (2006). The human infrastructure of cyberinfrastructure. *CSCW'06*, 483–492. https://doi.org/10.1145/1180875.1180950

Lin, Y.-W., Procter, R., Halfpenny, P., Voss, A., & Baird, K. (2007). An action-oriented ethnography of interdisciplinary social scientific work. In *Proceedings of the 3rd e-Social science conference, October 7–9*. Ann Arbor, MI: The University of Michigan. Retrieved from http://ess.si.umich.edu/papers/paper179.pdf

McNeese, M., Rentsch, J., Burnett, D., Pape, L., & Menard, D. (1998). *Testing the effects of team processes on team member schema similarity and team performance: Examination of the team member schema similarity model*. United States Air Force Cognitive Systems Division Report AFRL-HE-WP-TR-1998-0070, 45pp.

Nemeth, C., O'Connor, M., Klock, P. A., & Cook, R. (2006). Discovering healthcare cognition: The use of cognitive artifacts to reveal cognitive work. *Organizational Studies*, *2*, 1011–1035. https://doi.org/10.1177/0170840606065708

Olson, G. M., & Olson, J. S. (2000). Distance matters. *Human-Computer Interaction*, *15*(2), 139–178.

Orlikowski, W. J., & Yates, J. (2002). It's about time: Temporal structuring in organizations. *Organization Science*, *13*(6), 684–700.

Østerlund, C., & Carlile, P. (2005). Relations in practice. *The Information Society*, *21*(2), 91–107.

Pinch, T. J., & Bijker, W. E. (1987). The social construction of facts and artifacts: Or how the sociology of science and the sociology of technology might benefit each other. In W. E. Bijker, T. P. Hughes, & T. Pinch (Eds.), *The social construction of technological systems: New directions in the sociology and history of technology* (pp. 17–50). Cambridge, MA: MIT Press.

Pollock, N., & Williams, R. (2010). E-infrastructures: How do we know and understand them? Strategic ethnography and the biography of artefacts. *Computer Supported Cooperative Work (CSCW)*, *19*(6), 521–556.

Punch, K. F. (2013). *Introduction to social research: Quantitative and qualitative approaches*. Thousand Oaks, CA: Sage.

Ribes, D. (2014). Ethnography of scaling, or, how to a fit a national research infrastructure in the room. In *Proc CSCW* (pp. 158–170). New York: Association of Computing Machinery (ACM) Press.

Ribes, D., & Finholt, T. A. (2009). The long now of technology infrastructure: Articulating tensions in development. *Journal of the Association for Information Systems*, *10*(5), 5.

Ribes, D., & Lee, C. P. (2010). Sociotechnical studies of cyberinfrastructure and e-Research: Current themes and future trajectories. *JCSCW*, *19*(3–4), 231–244. https://doi.org/10.1007/s10606-010-9120-0

Ribes, D., & Polk, J. B. (2014). Flexibility relative to what? Change to research infrastructure. *Special Issue of Journal of the Association for Information Systems*, *15*, 287–305.

Russell, S., & Williams, R. (2002). Social shaping of technology: Frameworks, findings and implications for policy with glossary of social shaping concepts. In K. Sorensen & S. Russell (Eds.), *Shaping technology, guiding policy: Concepts, spaces and tools* (pp. 37–132). London: Edward Elgar.

Salas, E., & Fiore, S. M. (Eds.). (2004). *Team cognition: Understanding the factors that drive process and performance*. Washington, DC: American Psychological Association.

Sawyer, S. (2000). Packaged software: Implications of the differences from custom approaches to software development. *European Journal of Information Systems*, *9*(1), 47–58.

Sawyer, S., Crowston, K., & Wigand, R. T. (2014). Digital assemblages: Evidence and theorising from the computerisation of the US residential real estate industry. *New Technology, Work and Employment*, *29*(1), 40–56.

Sawyer, S., Østerlund, C., & Kaziunas, E. (2012). Social scientists and cyberinfrastructure: Insights from a document perspective. In *CSCW'12, February, 11–15*. New York: Association of Computing Machinery (ACM) Press.

Söderholm, H. M., Sonnenwald, D. H., Cairns, B., Manning, J. E., Welch, G., & Fuchs, H. (2008). Exploring the potential of video technologies for collaboration in emergency medical care: Part II. Task performance. *JASIST*, *59*(14), 2335–2349.

Stanton, N. (2016). Distributed situation awareness. *Theoretical Issues in Ergonomics Science*, *17*(1), 1–7.

Star, S. L. (2010). This is not a boundary object. *STHV*, *35*(5), 601–617.

Star, S. L., & Ruhleder, K. (1994). Steps towards an ecology of infrastructure: Complex problems in design and access for large-scale collaborative systems. *CSCW'94*, 253–264. https://doi.org/10.1145/192844.193021

Star, S. L., & Ruhleder, K. (1996). Steps toward an ecology of infrastructure: Design and access for large information spaces. *Information Systems Research*, *7*(1), 111–134.

Steinhardt, S. B., & Jackson, S. J. (2015). Anticipation work: Cultivating vision in collective practice. *Proc CSCW*, 443–453.

Vertesi, J. (2014). Seamful spaces: Heterogeneous infrastructures in interaction. *Science, Technology & Human Values*, *39*(2), 264–284.

Watson-Manheim, M. B., Chudoba, K. M., & Crowston, K. (2012). Perceived discontinuities and constructed continuities in virtual work. *Information Systems Journal*, *22*(1), 29–52.