Journal of the Association for Information Systems
JAIS

Special Issue

# Flexibility Relative to What? Change to Research Infrastructure

**David Ribes**
Georgetown University
dr273@georgetown.edu

**Jessica Beth Polk**
Cornell University
jess.b.polk@gmail.com

## Abstract

*It is often said that, in the face of an ever-changing world, infrastructure must remain flexible. Yet, what is meant by change remains glib and, consequently, so too do our studies on flexibility. In this paper, we develop three sensitizing concepts to investigate change to research infrastructure: 1) technoscientific: changes in research objects, scientific methods, and instruments; 2) sociotechnical: changes in social organization, coordination and, collaboration tools; and data sharing techniques; and 3) institutional: changes in funding and regulatory regimes. The majority of studies of "information infrastructure" have focused on the sociotechnical facet, and so we offer the two additional facets of change to help sensitize researchers to empirical instances of these encountered in the field, and to broaden the research agenda. To elaborate these concepts, we focus on a long-term research infrastructure that has been investigating HIV disease for nearly thirty years: The Multicenter AIDS Cohort Study (MACS). Over time, the MACS has faced tremendous changes in its science, collaboration and communication tools, its data and specimen repositories, its institutional environment, and the disease itself. Before we can begin to characterize flexibility, we must understand the nature of change research infrastructures face. We conclude by outlining a research agenda that will match forms of flexibility to the heterogeneity of changes an infrastructure may encounter.*

***Keywords**: Infrastructure, Long-Term, Sustainability, Change, Flexibility, Science, Ethnography and Archival Research, Symbolic Interactionism, Methodology.*

* Eric Monteiro was the accepting senior editor. This paper was submitted on 14th January 2013, and went through two revisions.

Volume 15, Special Issue, pp. 287-305, May 2014

# Flexibility Relative to What? Change to Research Infrastructure

## 1. Introduction

In this paper, we call for a broader inspection of change to research infrastructure so that, in addition to sociotechnical transformations in information technology, we may also inspect change to what we call the technoscientific and institutional facets of research infrastructure.

We examine the case of the Multicenter AIDS Cohort Study (MACS), a research infrastructure supporting the study of AIDS that was founded in 1983 and that has persisted for thirty years in the face of myriad changes, such as the dynamic nature of its primary research object, AIDS; its advancing scientific instruments; the reconfiguration of its social organization; the structure and regulation of its data; and how the study is funded. Throughout these transformations, the MACS has remained “the same study”: that is, it has continued to support the investigation of HIV disease and those who have it.

How do we conceptualize the types of changes that the MACS has faced over time? This paper offers three sensitizing concepts we call the facets of change: sociotechnical, technoscientific, and institutional. The sociotechnical facet is the intersection of social organization, coordination, and collaboration technologies, users, and information systems. The sociotechnical facet is most closely associated with the concept of information infrastructure, though we will show how they are not coextensive. We argue that, in recent studies of information infrastructure, it is the sociotechnical facet that has received the greatest empirical and conceptual attention. As a corrective to the narrow focus on the sociotechnical, we offer the additional concepts of the technoscientific facet—scientists and technical specialists, their instruments and methods of investigation, as well as their objects of research—and the institutional facet—the funding and regulatory environments that support and monitor infrastructure.

Ultimately, we define a more precise theoretical ground to revisit the family of concepts associated with “flexibility”, such as plasticity, resilience, and adaptability. The question “how should infrastructure adapt to a changing world?” is often answered with some form of the recommendation “remain flexible”. However, we feel that in studies of infrastructure, flexibility has been largely defined relative to the sociotechnical facet of change, and, consequently, most of our knowledge of flexibility has been drawn from investigations of that single facet. For example, Dourish and Edwards (2000) define flexibility as “the ability to accommodate individual difference in needs, use, style or task”. They note that an ideal quality of such systems is that they be “functional regardless of circumstance”. In this formulation, the source of change is the user, and flexibility is a property of the information system. In a similar vein, Pollock, Williams, and D’Adderio (2007) have called *generification* those design efforts that help prepare and plan for a technology’s adoption by heterogenous users and across differing contexts.

We do not contest these definitions and concepts, only their reach: we consider users a single source of change and sociotechnical system design and adaptation as one root of flexibility.

Accounting for the flexibility of research infrastructure in the face of change demands an inspection of features beyond the design of information infrastructure, systems, and practices. We find that change may be rooted in transformations to the practice and objects of science (technoscientific), and in funding and regulatory regimes (institutional). A robust and adaptable sociotechnical architecture is necessary but not sufficient to explain successful adaptations to change in each facet.

In this paper we inspect our case by way of the sensitizing concepts, generating a granular view of change to the MACS over time. We find that the MACS encountered changes that are within the purview of its role as infrastructure, but which cannot be described as sociotechnical alone. Rather than provide a novel definition of flexibility, we splinter its meaning as “flexibility relative to change”, and pose new questions for future studies. In the biography of a research infrastructure (c.f., Pollock & Williams, 2008), each facet presents a distinct problematic of change. Only after we understand change in its various forms can we begin to explore the breadth of the peculiar, future-oriented quality that is called “flexibility”.

## 2. What is Change? The Facets as Sensitizing Concepts

In a well-known methodological dictum by anthropologist (and cyberneticist) Gregory Bateson, he stated, “What can be studied is always a relationship or an infinite regress of relationships. Never a ‘thing’” (Bateson, 2000). Inspired by such relational thinking, Susan Leigh Star translated this insight into her definition of infrastructure: “infrastructure is a relation” (Star, 1999). By this assertion, she accomplished what has become a key move for infrastructure studies; that is, she relieved infrastructure of a narrow definition as “tubes and wires” or “heavy iron”. Instead, infrastructure has come to mean those resources and services—whether human, technical or sociotechnical—that enable, support, and shape activity. The housework of a stay at home mother, a classification system for books, and the taking of blood pressure by a nurse all became inspectable as infrastructure; that is, as a relation of support, and of resource and service provision.

Pursuing this standpoint, we offer the three facets of change—technoscientific, sociotechnical and institutional—to encourage a methodical inspection of infrastructure beyond a narrow definition as information technologies and their proximate human handlers.

The facets of change are called sensitizing concepts (Glaser, 1978) in the Symbolic Interactionist tradition of Sociology. We have developed these concepts to make sense of the MACS specifically, and research infrastructure more generally. These concepts are not intended to identify discrete social worlds or kinds of actors. It is not that there are one set of people doing science, another organizing collaboration, and yet another conducting science policy. In practice, many actors actively work across and align these facets. Furthermore, changes may occur simultaneously across the facets—they are not temporally or causally distinct.

Rather than identifying discrete social groups, these sensitizing concepts draw the analyst’s attention to the heterogeneous range of activities that are observable in the practice and processes of sustaining infrastructure. We broaden the research agenda of investigation in infrastructure studies, and offer a corrective to analyses that have become disciplinarily Balkanized.

Traditionally, different scholars of science, technology, and organization have focused on different facets of change to infrastructure and often treated them independently. For example, information studies has largely been concerned with information technologies, particularly communication and collaboration, data archiving, sharing, and interoperability (Hedstrom, 2001; Hine, 2006; Carlson & Anderson, 2007). Science and Technology Studies (STS) scholars have often focused on epistemic changes in fields (Kuhn, 1962; Bowker, 2006), the practice of science (Goodwin, 1995), and instrumentation (Keating & Cambrosio, 2003). Institutional and regulatory environments have been the concern of organizational and policy scholars (DiMaggio & Powell, 1991; David & Spence, 2003). But such changes are not discrete: they often occur concurrently and in an interlinked manner. An ecological approach to change is needed, bringing together multiple facets of change by investigating events and processes of transformation in and across the history of an infrastructure.

A caveat on the use of the term “information” in this paper: In each facet, we find issues that are usually associated with information, such as data sharing, communication and coordination tools, or the “invisible workers” of infrastructure such as data managers and technicians. However, we also find features beyond the (usual) formulation of information infrastructure that are necessary in order to track the biography our case study. Thus, while our analysis places materiality, technology, and technical practice at center stage, it is not limited to those technologies or practices that are soley “informational”. For example, as we see in Section 4.1.2, the development of an antiretroviral treatment for HIV disease in the mid-1990s was a landmark shift for the MACS research infrastructure: notably, treatments are a pharmaceutical technology rather than an informational one. In addition to focusing on changes, we also track continuities across the history of the MACS; only some of these are informational, such as data comparability. Other key scientific resources in the MACS, such as longitudinal blood sample archives, have a more orthogonal relationship to information, i.e., they are material specimens, and thus not reducible to information, however, it is information practices such as metadata creation that sustain their availability and meaning (e.g., the source of the specimen and when it was collected, are relations preserved as data). A study encompassing the three facets of change will necessarily intertwine, but also occasionally overflow, the concept of “information infrastructure”. We use the term “research infrastructure” to keep the purposeful and intentional orientation of this organization at center stage, that is, supporting science on AIDS.

## 2.1. Sociotechnical Change

Recent research has sought to approach infrastructure as sociotechnical by locating the enabling capacities of infrastructure at the meeting points of practical work, social organization, and technological arrangement. The term sociotechnical evokes a vision of collective human action in a material environment; coupled with the concept of infrastructure, sociotechnical also evokes a secondary order of often backgrounded machine and human organization (Ribes, Jackson, Geiger, Burton, & Finholt, 2012; Lee, Dourish, & Mark, 2006).

The earliest and most developed research program in sociotechnical studies of research infrastructure has concentrated on collaboration at a distance (Olson, Finholt, & Teasley, 2000; Olson & Olson, 2000; Olson, Zimmerman, & Bos, 2008). This research has investigated the dangers of geographic distribution (Kraut, Egido, & Galegher, 1988; Dourish & Bellotti, 1992; Fussell, Kraut, & Siegel, 2000; Nardi, Whittaker, & Bradner, 2000), the affordances of computer-mediated communication, and possible of advantages “beyond being there” (Hollan & Stornetta, 1992). A second well-developed line of sociotechnical research has examined the preservation and sharing of data (Birnholtz & Bietz, 2003; Zimmerman, 2008). Sharing scientific data is a central goal of research infrastructure, but data are often poorly curated and thus incomprehensible to others (Kellogg, Orlikowski, & Yates, 2006), or hoarded as scientific capital. Sociotechnical approaches to facilitate data sharing, such as ontologies (Ribes & Bowker, 2009) and metadata standards (Almklov, 2008; Zimmerman, 2008), are persistently difficult to implement.

Activity enabled by infrastructure can be cast as a case of “user meets system”: whether this is collaborating with geographically distributed colleagues or sharing data and findings. “Below” such first-order activities reside a network of human and technical actors maintaining and sustaining those operations, such as data managers and systems administrators (Karasti & Baker, 2010).

Correspondingly, in sociotechnical studies, the most common understanding (and emphasis) of change to infrastructure has also rested at the intersection of user and system. For investigators who focus on the sociotechnical facet, change is often presented as a matter of novel technologies, such as collaboration and coordination tools, that are adopted or enacted in ongoing organizational endeavors (Lawrence, 2006; Zimmerman & Finholt, 2007). Infrastructure scholars have studied sociotechnical transformations as a matter of changing information technology (Lawrence, 2006), forms of data (Borgman 2012), scale or extension of the network (Monteiro, 1997), everyday practice and work (De La Flor, Jirotka, Luff, Pybus, & Kirkham, 2010), social structure and organization (Lee et al., 2006) and temporalities, patterns, and rhythms of collaboration (Jackson, Ribes, & Buyuktur, 2011). These concerns have been most prominently articulated in studies on the sustainability of infrastructure (Ribes & Finholt, 2009; Bietz, Ferro, & Lee, 2012), and a smaller set of studies concerned with the often invisible work of maintenance and repair (Graham & Thrift, 2007; Jackson, Pompe, & Krieshok, 2012).

## 2.2. Technoscientific Change

However, research infrastructure is not only a sociotechnical system ready at hand, but also technoscientific, i.e., purposefully oriented to investigating an array of research objects (Ribes & Polk, 2012; Ribes, 2014). Over time, research infrastructure also faces changing objects of research, instruments, and methods of science.

We adopt the term “technoscience” from the work of sociologist Bruno Latour (1987). The term emphasizes the technically mediated quality between the doers of science and the objects of research. We call *objects of research* those things, processes, and relations that are of interest to scientists; these are often products of long chains of practical activity and technical intervention that make them available *as* objects (Latour, 1999). The "technical" in Latour's "technoscientific" includes both scientific instrumentation and practical technique, and forefronts that research is conducted and negotiated collectively between many participants such as scientists, technicians, and staff.

The field of Science and Technology Studies (STS) has developed methodological tools to analyze "science itself" as a part of social action in research infrastructure (Knorr-Cetina, 1999; Star, 1999; Beaulieu, 2010). STS has had a long focus on change to science (i.e., most famously Kuhn and paradigm revolutions), but this paper focuses on change to science as supported by infrastructure. For example, Paul Edwards (1999, 2010) has investigated climate science modeling infrastructures as a matter of collaboration and competition between meteorologists and climate scientists. Geoffrey Bowker and Susan Leigh Star (1999) studied the history of the International Classification of Disease as a boundary infrastructure that enables heterogeneous scientific experts to collaborate without consensus (Star & Griesemer, 1989). In these analyses, changes to information technology are coupled to changes in scientific methods and objects.

What is key here is that all studies of research infrastructure assume that its form has consequences for the doing of science. Regardless of whether this relationship is cast in a soft argument in which infrastructure speeds or slows knowledge production, or as a strong argument that sees the content of scientific knowledge itself as shaped by infrastructure, the central point is that infrastructure's design, use, and operation has consequences for scientific activity. This is why we study research infrastructure.

## 2.3. Institutional Change

Research infrastructure also has an institutional facet, whereby a significant investment in resources are intended to be temporally persistent at the scale of years to decades in a defined regulatory sphere (David & Spence, 2003; Mackie, 2007; Kee & Browning, 2010; Jackson, Steinhardt, & Buyuktur, 2013). In short, research infrastructures inhabit multiple overlapping institutional fields (DiMaggio & Powell, 1983; Hughes, 1999). They regulate scientific conduct and human subjects investigation; academic credit and rewards (Birnholtz, 2006); funding awards, evaluation, and renewal; data curation and public sharing (Borgman, 2012); and the employment of professors, scientists, graduate students, and staff.

"Institutional" here does not refer to a distinct sphere or a macro-scale of analysis. Drawing on the concept of institutional work (Lawrence, Suddaby, & Leca, 2011), this sensitizing concept should also remind us of actors' practical orientation as they engage and re-engage with institutional concerns. Managing institutional change is work done with an eye to regenerating funding and bridging the constant drift between emerging policy and how the infrastructure is run on a day-to-day basis—activities with an orientation to what Ribes and Finholt (2009) call "the long now of infrastructure".

The institutional facet is not a "stable backdrop" for research infrastructure, i.e., the national agencies of science are themselves in motion. For instance, in the U.S. beginning in the early 2000's, a systematic emphasis on the development of research infrastructure for the sciences was made at the level of national funding agencies. Many projects today dubbed "cyberinfrastructure" (the term of art for information technology-driven research infrastructure) were funded following the release of the influential *Revolutionizing Science and Engineering Through Cyberinfrastructure* report (Atkins, 2003), a policy document describing and advocating large-scale, long-term, community-wide research infrastructure. Many of these cyberinfrastructure projects were funded by the National Science Foundation (NSF) —the state funding agency in the US—through particular grant programs such as information technology research (ITR) or knowledge discovery initiative (KDI) (Cummings & Kiesler 2003, 2005, 2007; Cummings, Finholt, Foster, Kesselman, & Lawrence, 2008). As such, they were subject to peer review and to the stipulations of award contracts (Lamont, 2009). Such infrastructures are the object of regimes that regulate the production of data and how those data are made available to other scientists and the public (Vertesi & Dourish, 2011).

Speaking more generally, agencies regularly discontinue past modes of funding as they create new avenues for support. The criteria for evaluating ongoing projects are revised, as are the processes that lead to renewal, budget increases, and slashes, or the “sunsetting” of entire programs. Most notably, in the past thirty years, policy for data ownership has been significantly reshaped, increasingly emphasizing sharing and reuse. Borgman (2012) explores the tensions between everyday data practices of scientists and the data sharing policies of agencies such as NSF. Howison and Herbsleb (2011) investigate comparable issues in the sharing of scientific software. They note that funding agencies, systems of academic reward and credit (an institutional issue), and relations with industry affect how and when sharing occurs. All of these institutional changes occur in what Karasti, Baker, and Millerand (2010) call “infrastructure time”—a timeframe longer than a grant or one-off study, often even longer than individual careers.

Each of the three facets is in motion, subject to particular regimes of change over time. Just as the tools and techniques of collaboration are changing (sociotechnical), so too is the science itself, with new objects of research, fields of inquiry, emerging specializations, and instruments (technoscientific). Funding and regulatory arrangements for research are not a stable backdrop either: they inevitably shift as the agencies supporting science enact annual budgets or alter rules for data sharing (institutional).

## 3. The MACS, the Visit, and its Archive

This paper examines change in and across the history of our case study: the Multicenter AIDS Cohort Study (MACS). Launched in 1983 in response to an emerging epidemic, MACS initially focused on understanding the cause and patterns of transmission for what was, at the time, known as Acquired Immune Deficiency Syndrome (AIDS)[1]. To do so, the MACS recruited a cohort of gay and bisexual men from four American cities — a group of subjects the MACS continues to follow to this day. This population was chosen based on their “at-high-risk” status for manifesting symptoms of the syndrome. Notably, the MACS was founded before the infectious agent of AIDS was officially identified in 1984, which received its current name in 1986: the Human Immunodeficiency Virus (HIV). This is our first example of a technoscientific change, which we elaborate in Section 4.1.1. MACS’ research program has since expanded to investigate the natural and treated history of HIV, along with the efficaciousness and adverse effects of treatment over time.

The MACS is geographically distributed (hence, “multicenter”) across four American cities: Baltimore/Washington, Chicago, Los Angeles, and Pittsburgh. Each of the four sites has a principal investigator (PI) as the scientific and administrative lead; an additional PI was added in 1987 to help manage data and to facilitate statistical analysis (described in detail in Section 4.2.1 as a sociotechnical transformation). The MACS is funded through the National Institutes of Health (NIH), primarily through the National Institute of Allergy and Infectious Diseases (NIAID), but in lesser amounts through other agencies such as the National Cancer Institute (NCI).

The MACS’s central mandate is to continue to generate data and specimens (thus, a prospective rather than solely retrospective study), and to make these resources available for scientists to use in their research. This is the most direct meaning of the MACS as a research infrastructure. As we see below, a great deal has changed in the MACS’s history, but, across these changes, it has remained “the same study” as members have sought to regenerate the continuity of its past, present, and future data and specimen archive.

While we focus on change in this paper, the concept of the technoscientific facet has sensitized us to two continuities across the study’s history: the visit, and the data and specimen archive. “The visit” is what MACS members call the biannual trips that participating men (the subjects) take to a MACS-affiliated clinic in one of four cities. In 2013, all four MACS sites conducted visits 58 and 59. All data and specimens are collected from these cohorts, and it is at the visit that the majority of these collections occur. During this visit, the participating men provide a new batch of specimens and spend time completing questionnaires. In other words, the visit is the moment when subjects are translated into data and specimens by the use of instrumentation such as questionnaires, techniques such as blood pressure readings, and interventions such as blood taking.

[1] Both the disease and the virus have had changing names and meanings throughout the years (Epstein, 1996). For the sake of clarity and simplicity, we use the terms AIDS and HIV in this paper.

In many senses, the men and their biannual visit is the study's persistent kernel[2]: the source of the comparable, standardized dataset and specimen collections that makes the MACS the same ongoing longitudinal study since its inception. Neither the specific data and specimens nor how they are made available for scientific use has remained the same over time—it is the availability and comparability of these data and specimens that the MACS has sought to regenerate across technoscientific, sociotechnical, and institutional changes (Ribes, 2014).

Our study of the MACS is archival, ethnographic, and interview and action-research based. Our primary data source has been the extensive documentary archive generated by the study's members over the past thirty years, which include published papers, formal reports, and oral history interviews. Additionally, the second author worked at the MACS's Baltimore/DC site for two years. The study's members and its subjects generously permitted us to conduct interviews and focus groups. Both authors have collaborated with the MACS to investigate online health information-seeking behaviors among its D.C. cohort. Our data collection and analysis method was largely informed by grounded theory (Glaser & Strauss, 1973), and the "follow the actors" approach of actor-network theory (Latour, 1987). Our sensitizing categories were generated by iterative returns to field sites, document archives, and members.

## 4. The MACS and the Facets of Change

As noted above, the facets of change are sensitizing concepts used to discover empirical instances of change in the study of a specific infrastructure. As Herbert Blumer (1954, p. 8) notes: "what the concept refers to is expressed in a distinctive manner in each empirical instance and can be got at only by accepting and working through the distinctive expression". In other words, in the abstract, the facets of change are sweeping notions without specific referents; it is only as they are worked through an empirical instance that they gain a concrete meaning.

In Sections 4.1 to 4.3 we do just that by tracking five instances of change in the study's history: 1) transformations in the research objects and subjects, 2) changes in scientific specializations and instrumentation, 3) changes in the data architecture and organization, 4) the regulatory regime for data, and 5) funding mechanisms. The changes themselves operate across the facets, but, for each change, we emphasize its relationship to one facet.

### 4.1. Technoscientific Change

To assert that research infrastructure has a technoscientific facet is to be analytically insistent about the practices of science an infrastructure is supporting. Research infrastructures are assemblages of people, instruments, and techniques oriented to the investigation of particular objects of research—for instance, the earth (Ribes & Bowker, 2008), climate (Edwards, 2010) and the brain (Lee et al., 2006), or the public (Igo, 2007). Such objects and their methods of investigation may themselves change over time.

We define technoscientific change as the shifts in science that comprise: 1) research objects: the things, processes, or relations of interest to the scientific inquiry; 2) methods and instrumentation: the techniques and technologies for investigating research objects and for generating data; and 3) researchers: the scientists inhabiting (inter)disciplinary communities oriented to phenomena and who carry out investigations in practice.

In examining the MACS, we see how transformations in AIDS—the organization's central research object—directly led to reorientations in the study's program.

#### *4.1.1. From AIDS to HIV: Emergent Objects of Investigation*

The MACS was initiated in 1983 to begin making sense of the causes and modes of transmission of a potential epidemic that the medical community had only recently agreed to call AIDS. At the time, many causes were posited: an infectious agent was considered likely in medical circles, but

[2] The paper (Ribes, 2014) defines the concept of the kernel, and outlines four resources and services the MACS has sought to renew over time. In this paper, we focus only two of these resources and services.

investigators also prepared for explanations such as use of amyl nitrate (“poppers”—a chemical inhalant popular in gay nightlife at the time), environment or diet, and several other possibilities (Epstein, 1996). AIDS was statistically linked to gay men, but not yet causally to sexual behavior.

To conduct this study, MACS investigators recruited 4, 954 gay and bisexual men considered to be “at risk” for manifesting symptoms. MACS epidemiologists would track these men to understand AIDS’s natural history. These participants were asked to come to clinics twice a year to fill out questionnaires and provide biological specimens (i.e., the visit).

In 1984, HIV was announced as the retroviral agent that leads to AIDS. A scientific study tasked with finding an agent suddenly found itself with an answer. From a speculative causal landscape, the MACS was now operating firmly in the domain of virology and sexually transmitted diseases.

Did finding a cause mean that this the study was no longer needed? Clearly not. The identification of a new scientific entity heralds a new object of investigation: discovery of a retroviral agent thereafter revised many research topics such as infectiousness, natural history, treatment, and the possibility of a cure. The MACS would have to reorient itself to make sense of this novel entity.

How did this distinct change in the purpose of the MACS and of its central research object impact the visit and its archives of data and specimens? Surprisingly, very little. Because in 1983 the causal agent was unknown, the study’s strategy in its earliest phases had been to collect and store just about every specimen they could think of: hair, blood, semen, stools, fingernails, and so on. Each of these materials could potentially serve as key indicators for the causes or modes of transmission for the disease. Similarly, they collected a broad swath of behavioral data: sex partners, frequency, acts, and locations; drug use; medical history; family health; and so on. The data and specimen archive was assembled to include environmental, biological, and systemic etiologies conceivable at the time.

As HIV was confirmed as the agent, the MACS was—at the intersection of science, fortune, and preparation—equipped for this change in the landscape of scientific knowledge. The wealth of specimens and data had “future proofed” the study to this new viral entity. In the years following 1984, the scientific community would come to a broad agreement about the key carriers of the disease—such as blood—and the key activities associated with its transmission—such as receptive anal intercourse; the MACS was already collecting these specimens and behavioral data. Thus, immediately following this announcement, we find no new specimen collections or questionnaires administered to MACS participants during the visit.

HIV’s discovery is an exemplary technoscientific change for research infrastructure: a new entity, HIV, was placed at the very center of this research enterprise.

#### *4.1.2. From Fatal to Chronic: Expanding Experts, Methods, and Instruments*

HIV’s discovery came as an early change for the MACS. Since then, the flood of technoscientific changes has continued unabated. This section focuses on how the MACS adapted its science as pharmaceutical treatment transformed HIV infection into a chronic disease. This change required a very different response from the MACS than HIV’s discovery.

In 1987, azidothymidine (AZT) became available as the first approved treatment for AIDS; however, it was not until 1996 that highly active antiretroviral therapy (HAART, now simply ART, and popularly known as “the cocktail”) rendered HIV infection into a manageable illness. Thereafter, for many, what had been an acute and terminal disease (often referred to as “a death sentence”) in the 1980’s and early 1990’s became a manageable and chronic condition.

Dying from a disease is a monumentally different experience than living a relatively healthy life with a manageable chronic condition. In turn, studying a population of increasingly ill subjects is quite a different research endeavor than studying subjects living into their twilight years. From the perspective of MACS investigators, following the availability of the treatment the same cohort could now be tracked for decades rather than years.

On the one hand, treatment had prolonged life, but, on the other hand, those pharmaceuticals presented their own complications. For example, the toxicity of the treatment—such as its effects on the liver—became a new research object for the MACS. Furthermore, the intersection of HIV infection, treatment, and, eventually, aging presented completely unique complications. New objects of research—such as liver disease or blood pressure—required adding new specialists with their own unique set of scientific methods, questions, and instruments (e.g., gastroenterologists, hepatologists) to the MACS. In the case of HIV and liver disease, for instance, MACS adopted FibroScan techniques as a means of measuring fibrosis of the liver (or cirrhosis) via ultrasound technology (a non-invasive alternative to performing a physical liver biopsy).

While in the first example of technoscientific change—HIV's discovery—the activities of the visit, and namely the data and specimen collection, went on largely unchanged, with HAART treatment and the transformation of the disease into a chronic condition, the visit and data collection had to be expanded with new kinds of data and far longer longitudinal spans for the visit. Here, it was the subjects themselves who changed, generally living longer and healthier lives but with all the complications associated with managing a chronic disease. To continue its study of this disease, the MACS found that it needed to change its expert base and its data and specimens to study a chronic disease and the effectiveness or toxicity of treatment.

In this section, we characterized two distinct technoscientific changes in the history of the MACS. Aside from the characterization that these changes occurred in the technoscientific facet, there is nothing similar in the "kind" of change that occurred. The first change—the discovery of HIV—shifted the study's purpose from seeking causes to understanding a new causal entity and its disease trajectory. This is very different from the availability of treatment, which transformed the nature of the disease itself and the lives of those who had it.

## 4.2. Sociotechnical Change

The technoscientific facet is largely about the "what" of research infrastructure; it denotes what has been collected and preserved for later scientific use (specimens and data), and the "what" that is being investigated: an unknown cause, a fatal disease, or a chronic illness. In contrast, the sociotechnical facet is a great deal about the "how" of making materials, expertise or instruments available *as* resources and services (i.e., how data are preserved and how they are shared). It is in this facet that we will most centrally find what is most often referred to as information infrastructure—the technologies and techniques for manipulating, preserving, representing, and disseminating data, and for supporting temporally and spatially distributed coordination and communication.

The sociotechnical facet of research infrastructure is transformed through the development of new 1) organizational forms, 2) methods of collaboration, and 3) techniques of data and specimen sharing. In this section, as an example of sociotechnical change, we examine the founding of a data center in 1987, four years after the MACS study was initially funded. To the original four principal investigators (PIs), an additional PI was appointed, who was tasked with managing and curating data across the four sites, and with providing (bio)statistical methodological expertise. The downstream consequences of this punctuated moment of change reshaped the social organization of the MACS, how participants collaborated, and how data came to be shared.

### *4.2.1. Bringing Data Home*

From 1983 to 1987, the MACS's data were housed at the NIH campus in Bethesda, Maryland. As the study gained momentum (i.e., new instruments and research specializations) and more data were generated at each of the four sites, MACS scientists reported problems getting data to the repository at NIH, and, more importantly, getting data out. In particular, one PI noted that "making cross-site comparisons was very difficult, the very purpose of our study!" (interview). In short, the central mandate of the MACS—making data and specimens available for use in cross-site studies—was not being fulfilled.

In response, NIH issued a request for proposals and competed a new MACS site, specifically charged with managing and analyzing data. In 1987, the award went to John Hopkins University, and was called CAMACS (Center for the Analysis and Management of the Multicenter AIDS Cohort Study).

Organizationally, CAMACS created new roles (i.e., a PI and a (bio)statistically trained staff), and these individuals in turn introduced new organizational standards for data collection, management, and sharing at each of the four MACS sites. The introduction of CAMACS also resulted in changes to data collection itself: the ways in which specimens were obtained and questionnaires administered to the men participating in the study.

In discussing the initial transition from the NIH to CAMACS in 1987, one member of the CAMACS site recounts:

> *we made many trips down to NIH to transition the data to us … for example, we had to develop a whole system for code books and data....they* [the NIH] *had forms, they had data files, but no, there were no code books* (interview).

In 1987, part of the data were already computerized in flat-files, but another part of the data resided in paper forms. For example, the qualitative answers of participants and ad hoc pencil markings recorded by MACS clinicians: "they wrote things on the forms, so we had the forms" (interview). But no standardized protocol or metadata codes existed so that the meaning of data could be sustained or easily shared across sites and among MACS researchers. With the introduction of a codebook, CAMACS could document, archive, and standardize protocols, such as, the questions asked during a participant study visit.

The creation of a data management center was a very different kind of change on the part of MACS than the technoscientific changes described in Section 4.1. Rather than changes to the objects and subjects of the MACS, we find transformations in the daily practice of data management at each site, in the procedures to circulate data across the sites, and, more subtly, in the authority and control over the sharing and reuse of data.

Most contentious was the issue of control. The four original PIs now had to collaborate with a new, equal partner, a biostatistician with his own research specialty and a formal authority over the curation of data. For scientists, data are the raw material for generating insight and shoring up claims. Multiple MACS scientists have reported a concern that the new data center would have privileged access to the data from all four sites across the US. One PI said: "The concern ... was that the data center was going to take the data... and do all the work and analyze and do all the publication" (interview).

This could mean that CAMACS would be in a position to generate the most important findings: cross-site comparisons. In the end, the MACS PIs negotiated a compromise in which, roughly speaking, scientific findings would be first-authored by biological-medical scientists, and methodological findings would be first-authored by CAMACS scientists. This new mode of collaboration was a sociotechnical achievement between nationally distributed, heterogeneous disciplinary specialists, operating through a new regime of mediated data sharing.

Later, as HIV became chronic (a technoscientific "event" described in Section 4.1.2), it also transformed key organizational goals of the infrastructure (sociotechnical features). The MACS was always a longitudinal study, but it was not always long-term; rather, it became long-term as its research objects became lasting phenomena. As participants lived longer (while, simultaneously, HIV resisted efforts to find a cure), so too did the MACS: the study transitioned from an initial four-year contract to an indefinitely long-term research study equipped with the resources to track participants over decades. As we see in Section 4.3, as novel data sharing regulations were institutionalized, the distributed data flows and codebooks facilitated the MACS's entry into the emerging era of regulated scientific data sharing and potentially renewable funding.

### 4.3. Institutional Change

The institutional facet refers primarily to the funding and regulatory environment of research infrastructure. As with the technoscientific and sociotechnical facets, “institutional” is a sensitizing concept that draws the analyst’s attention to activities linked to infrastructure and, in particular, to efforts that seek to sustain infrastructure within a broader ecology of agencies and organizations that, for example, enact regulation and the distribution of funds. While the understanding of this facet in this paper is informed by neo-institutionalist theory (DiMaggio & Powell, 1983), we do not consider institutional to refer to “another scale” akin to macro-organizational analysis. We draw from Lawrence et al.’s (2011) concept of institutional work that emphasizes how “individuals actively engage in processes of institutional creation, maintenance, disruption, and change” (p. 53).

Some institutional actors are “outside” of the infrastructure, such as program officers in the agencies that fund science (in the case of the MACS, the National Institutes of Health (NIH)) and legislators or policy makers in state agencies. However, there are also many actors “in” the infrastructure that engage in institutional work. The principal investigators in the MACS exemplify such activity. They are practicing scientists who actively contribute to research in their fields. But they are also the leading actors in creating social order, scheduling regular meetings, and encouraging others to adopt collaboration tools. Notably, in founding the MACS, certain PIs did not simply respond to a funding request by NIH, but actively encouraged the allocation of funding and helped craft the request for proposals (RFP). This is not unusual. In America, it is a common misconception that funding agencies are distinct from science, however, this is not the case, e.g., program officers are usually drawn from the ranks of science, and in turn, they often return to those ranks after a stint within agencies; while they are in office they will regularly consult with their scientific constituencies to shape new programs.

As with the other facets, information infrastructure is implicated in institutional change, but not coextensive with it. Policy may regulate, for example, how the personally identifiable data of subjects must be securely stored, or what level of consent is required to share those data—these are information issues, broadly construed. Information managers are concerned with developing repositories that comply with emerging data security, privacy, and sharing policies. But questions of funding and regulation do not necessarily tie directly to the informational aspects of research infrastructure—for instance, in Section 4.3.2 we track how changing funding arrangements made the biannual visit to clinics more central to the infrastructure, while shifting the responsibility of funding research to individual investigators. We break down change in the institutional facet into two categories and for each offer an illustrative vignette: 1) regulatory and 2) funding.

#### *4.3.1. Regulatory: From MACS Data to Public Data Sets*

The 1990s witnessed a sea change in the regulatory environment for the MACS’s data archives. New attitudes towards scientific data sharing had been percolating for years in the US; rationales such as data reuse across studies or ease of replication had come to the fore (Borgman, 2012). These new data rationales critiqued models in which data were tacitly the possession of the collecting scientists. In 1995, under the directorship of Carl Dieffenbach, the Division of AIDS (DAIDS) in the National Institute of Allergy and Infectious Diseases (NIAID) adopted a new policy for data sharing, particularly for projects with budgets over $500,000 annually. Such regulations received a higher order of recognition and enforcement when, in 2003, a similar set of rules were mandated NIH-wide[3].

MACS scientists were not unreservedly supportive of the new policies. There are good reasons for scientists to be wary of sharing their data. Making data publically available is often perceived as threatening the epistemic authority and fruitfulness of the scientist who must “give it up”: public data may be reanalyzed, possibly revealing flaws in the analytic method, and scientists may fear “being scooped” if other researchers are able to generate findings from their own data first. Furthermore, making data sharable is an arduous and unrewarding task: data are collected in ways that make them undecipherable to outsiders, organized in ways idiosyncratic to those who collected them, or stored in formats that are not easily transferred. Making data easily shared—interoperable—is laborious and expensive work. And yet, the work of making data that can be shared rarely generates career rewards

[3] http://grants.nih.gov/grants/policy/data_sharing/

(e.g., contributing to tenure or generating peer recognition), and is to this day a largely unfunded mandate requiring significant technical labor.

This said, by the late-90s, the MACS was in good shape to address the new data-sharing regime. As one NIH program officer describes retrospectively:

> *The one thing about the MACS is they actually were the earliest adopters. Well, there may have been some initial resistance. I think that what they did is, within a very short period of time, saw the value of* [data sharing] (interview).

As we outline in Section 4.2.1, the MACS had already instituted a codebook (i.e., an approach to metadata) over a decade before data sharing become regulated by NIAID. These past changes had prepared the study for the new data regime. In this sense, the study was already largely able to manage these emerging institutional demands. Correspondingly, scientists had already been entrained in a normative data-sharing regime. The norm was to send new data to CAMACS, which, in turn, could distribute data to MACS sites.

That said, data did not simply "become public"; rather, MACS instituted a procedure for requesting and approving new instances of data sharing. Today, the MACS's executive committee (EC)—comprising PIs, NIH program officers, and study subjects—deliberates on each new request for data, and thus, act as gatekeepers that approve or deny a new sub-study: "namely, because it's either a core study of the MACS investigator or it's being done by some other investigators" (interview) among many other reasons.

In turn, CAMACS staff attempt to ensure that any new data produced through sub-studies be submitted back to the common data archive. In this manner, principal investigators maintain some control over their data, while still remaining flexible to new requests; and, in turn, new collaborators must adapt to the policies of the MACS that require standardized data submission and project updates.

### 4.3.2. Funding: From Contract to Cooperative Agreement

While infrastructure is intended to last, funding does not come with the same promise. Most nationally funded research programs in the US undergo regular evaluation and must apply (or "recompete") for renewal. In preparation for such renewals, principal investigators are prospectively considering a future of funding possibilities; they are aware of a definitive end-date to current funding and seek to consider how their current research efforts will be perceived and reviewed in the future by their peers and by funding bodies.

The MACS's core funding award from the National Institutes of Health (NIH) has been renewed seven times. At each occasion of renewal, the PIs and NIH program officers recount debates at the NIH, and within the broader scientific community, over whether to renew MACS funding or simply "sunset" its expensive prospective efforts (ongoing data collection through the visit) and make the archive of data and specimens available as a retrospective collection:

> *What people don't realize is, again, it* [MACS] *is this infrastructure, it's keeping a group coherent and working together as a group that takes money … So, we had all the data and the specimens, what are you going to do, let 200 different individual investigators, none of them* [who] *know the study very well, come in and propose to do things? … Or are you going to let the people who know the data extremely well continue to mine* [the data] *and use it to the maximum advantage*? And that is the argument that has won the day each time. (Interview)

Thus, due to the nature of the MACS's funding arrangement, the study has to embrace scientific change (or even seek it out) in order to remain not only relevant as a research enterprise, but also to continue to be considered a worthy financial investment. At each renewal, the balance of funding is renegotiated, as are the weightings of research objects for the study: aging, cancers, heart disease, new infections, etc.

A notable change in funding occurred as the MACS shifted from a “contract” to a “cooperative agreement”. In NIH, these categories refer to specific funding and administrative arrangements:

> *a contract, unlike a grant,* [.] *is a set of specifications and work that the NIH wants done, and asks specifically for proposals to do* [those] *and therefore is under much greater restriction and scrutiny in terms of the deliverables, the timeline, the specifications, etc … The NIH controls the work* (interview).

A contract leaves much of the administrative authority in the hands of the funding agency, while a cooperative agreement is closer to a traditional scientific research award in that it gives greater autonomy to the principal investigators.

The shift to a cooperative agreement also focused the MACS’s budget. Whereas the budget for the contract had included *both* core activities (i.e., specimen and data collection) *and* basic research, the cooperative agreement was primarily intended to support only core activities. Thereafter, the majority of basic research would be funded from additional competitively awarded research grants that would “attach” themselves to the study. This change marks the transition from what Ribes and Finholt (2009) call project to a facility: from a short-term venture with specified goals and a limited funding timeframe to a long-term support structure for ongoing science.

## 4.4. From “Sensitizing Concepts” to “Categories of Change”

To speak of technoscientific, sociotechnical, or institutional change is simply too broad to concretely characterize the MACS. Instead, these concepts are intended to orient the investigator to features encountered during empirical research. On their own, sensitizing concepts have a “low empirical content” (Kelle, 2007); that is, in the abstract, they do not have specific referents. It is only as they are brought to bear in guiding an empirical investigation that they begin to gain analytic value and concrete meaning (Blumer, 1954):

> *Sensitizing concepts can fulfill an important role in empirical research, since their lack of empirical content permits researchers to apply them to a wide array of phenomena. Regardless of how empirically contentless and vague they are, they may serve as heuristic devices for the construction of empirically grounded categories.* (Kelle, 2007, p. 208)

By inspecting the MACS through the facets of change, we have generated more granular categories of change this study has experienced over time. We summarize the categories we inspect in this paper below, and illustrate their relationship to the facets of change in Figure 1.

> *Objects and subjects:* Investigators continuously turn their attention to new objects of investigation, losing interest in other objects, and finding old objects recurring in surprising ways. Similarly, the subjects (participants) too are changing as they live their lives, age, or become more or less ill.
>
> *Methodologies and instruments:* The instruments and methods of science are in motion; new tools and techniques are adopted and old instruments retired. This includes social scientific and epidemiological techniques such as questionnaires and surveys, or biomedical instruments such as assays.
>
> *Researchers:* There is a shifting centrality of fields over time. The disciplinary training of investigators varies as some specializations become less relevant to objects of study and other specializations must be added to make new objects researchable. The patterns and forms of disciplinary collaboration too are reconfigured.
>
> *Data and specimens*: Generally, the data and specimen archives are growing, though specific

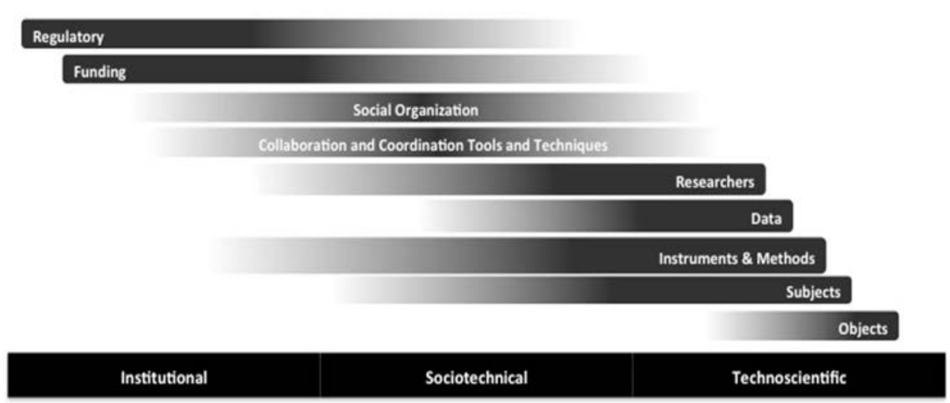
data and specimens are no longer collected. The techniques and technologies of data stewarding—preservation, curation, and sharing—are evolving, but so too are the challenges of scale and time. The responsibilities for this work shifts across institutions and individual actors, reshaping access to these key resources.

*Collaboration and coordination:* Spatial and temporal relations are markedly transformed by the introduction of new methods and tools for collaboration across disciplines and geographic distribution. Such tools do more than "collapse space and time", they also create new roles and agencies for the actors that introduce and sustain communication and coordination (Ribes, Jackson, Geiger, Burton, & Finholt, 2012).

*Social organization:* The organization may grow, or shrink, and in many ways: e.g., in its number of researchers, or its physical sites. New organizational responsabilities may also demand broad shifts in administrative or technical work, such as with the case of managing data curation and sharing. At a more granular level, there will be a turn-over in leadership, management and staff, and entirely new roles will be created or withdrawn.

*Funding and regulatory environments:* Mechanisms for funding, evaluation, and renewal are regularly revised, sometimes shifting the entire endeavor from one budgetary category to another. The justification of intellectual merit and broader impacts must be continuously attuned to the disciplines this infrastructure supports. Broadly speaking, policies regarding publication of results, and sharing data have increasingly emphasized openness, but always with new ways of regulating exchange.

The list above is provisional and specific, recounting what we found in applying the facets of change to the MACS, rather than a generalizable set of categories. Some of these categories, such as data, will be general to many research infrastructures. Other categories, such as human subjects, while absolutely central to the MACS, may be altogether absent from other cases, such as organizations supporting the ecological sciences (Ribes, 2014). Thus, the facets of infrastructure are analytic tools used to assemble specific categories of change for concrete cases.




This diagram illustrates the make-up of the three overarching facets of change as applied to the MACS. By overlaying our more granular categories of change, we can see that technoscientific, sociotechnical, and institutional changes are by no means discrete.

**Figure 1. The Three Overarching Facets of Change as Applied to the MACS**

## 5. Revisiting Flexibility: An Agenda for Research

In this paper, we begin to define a holistic vision of change to research infrastructure. We contribute to the literature by specifying three concepts that will sensitize the investigator to multiple forms of change. Our more distal goal is to lay the foundation that will be required to speak meaningfully of systems, organizations, or infrastructure that are responsive and resilient to sociotechnical, technoscientific *and* institutional changes. Here, we return anew to the family of concepts associated with "flexibility", such as adaptability, robustness, and extensibility. Having outlined the heterogeneous forms of change infrastructure may encounter, we conclude by articulating a future agenda for the investigation that seeks to understand infrastructure's ability to continue supporting activities in the face of such changes.

A next step, then, is to elaborate flexibility with respect to technoscientific, sociotechnical, and institutional change. Each facet requires a specific articulation: capacities and qualities of flexibility relative to change, and how they are achieved. This paper presents the analytic tools to conceptualize change—what an infrastructure must be "flexible relative to"—but it also hints at multiple strategies to achieve flexibility. Here, we briefly focus on technoscientific flexibility, the form of adaptability and resilience that is perhaps most central to research infrastructure in that it must support the doing of science across dramatic transformations to objects of research.

Across each form of change we have inspected, MACS has "gone on" supporting scientific investigations—that is, it continues to be a research infrastructure. Recall that MACS was founded in 1983 before the identification of HIV. Yet, following this monumental shift in the research program, the core collection activities (specimens and data) and subject base remained largely unchanged. In the face of a new ontological entity—what has come to be the very center point of the study—the MACS, its collection ritual, its four sites, and its disciplinary matrix of scientists continued much as it always had.

The MACS's ability to adapt to these changes presents a distinct form of technoscientific flexibility: techniques, technologies, and organizational innovations to adapt to changes in scientific objects, instruments, and methods of investigation. We argue elsewhere (Ribes, 2014) that the MACS's resilience and plasticity to this form of change is rooted in the technoscientific feature we call "the kernel"; specifically, its broad specimen collection routine, and the specific instruments (such as surveys) that transformed the "behaviors" of the participants into inspectable, and eventually archival, data. No design feature that we describe in the other two facets can explain the MACS's resilience to this change: no information technology and no institutional arrangement can supplement the specimens and behavioral data that allowed the MACS to go on in its research following HIV's discovery.

This is an altogether distinct form of flexibility than we would expect from an organization adapting to novel information technologies, for example, a new data management approach (c.f., Millerand, Ribes, Baker, & Bowker, 2013). Accounting for the flexibility of research infrastructure in the face of change demands the inspection of features beyond the design of information infrastructure, systems, and practices. Following from the elaboration of the three sensitizing concepts—technoscientific, sociotechnical, and institutional change—we believe that it will be necessary to proliferate kinds of flexibility. Or, at least, that an explanation of flexibility will need to account for these broader forms of change.

Whether we are discussing technological design, social organization, or their intersections, flexibility is not a single thing. Rather, flexibility is a capacity that can only be characterized relative to instances of change. In the past—by focusing on sociotechnical change to the exclusion of the specific research objects or institutional activities—we have been in danger of ignoring the importance of the circumstances and the particularities of change that infrastructure must be flexible toward. Just as we have come to understand infrastructure not as a "thing" or bounded entity but as a relation, so too should we consider flexibility as relative to forms of change.

## Acknowledgements

We thank the MACS members and participants for their generous contributions to this study, in particular Dr. Michael Plankey. This paper was presented at Syracuse University's School of Information Studies: we would like to thank the attendees at that event for their feedback, and in particular Steve Sawyer and Carsten Oesterlund. This research has been supported by NSF/VOSS award #0904145 "Delegating Organizational Work to Virtual Organization Technologies" and #1322275 "Flexible Research Infrastructure: A Comparative Study".

## References


Almklov, P. G. (2008). Standardized data and singular situations. *Social Studies of Science, 38*(6), 873-897.

Atkins, D. E. C. (2003). *Revolutionizing science and engineering through cyberinfrastructure: Report of the National Science Foundation blue-ribbon advisory panel on cyberinfrastructure*. Washington, DC: National Science Foundation.

Bateson, G. (2000). *Steps towards an ecology of mind*. Chicago, University of Chicago Press.

Beaulieu, A. (2010). From co-location to co-presence: Shifts in the use of ethnography for the study of knowledge. *Social Studies of Science, 40*(3), 453-470.

Bietz, M. J., Ferro, T., & Lee, C. P. (2012). *Sustaining the development of cyberinfrastructure: An organization adapting to change. Proceedings of the ACM Conference on Computer Supported Cooperative Work*, *Washington, USA,* 901-910. DOI: 10.1145/2145204.2145339

Birnholtz, J. P. (2006). What does it mean to be an author? The intersection of credit, contribution, and collaboration in science. *Journal of the American Society for Information Science and Technology, 57*(13), 1758-1770.

Birnholtz, J. P., & Bietz, M. J. (2003). Data at work: Supporting sharing in science and engineering. *Proceedings of the International ACM SIGGROUP Conference on Supporting Group Work, Sanibel Island, Florida, USA*, 339-348.

Blumer, H. (1954). What is wrong with social theory? *American Sociological Review, 19*(1), 3-10.

Borgman, C. L. (2012). The conundrum of sharing research data. Journal *of the American Society for Information Science and Technology, 63*(6), 1059-1078.

Bowker, G. C. (2006). *Memory practices in the sciences*. Cambridge, MA: MIT Press.

Bowker, G. C., & Star, S. L. (1999). *Sorting things out: Classification and its consequences*. Cambridge, MA: MIT Press.

Carlson, S., & Anderson, B. (2007). What are data? The many kinds of data and their implications for data re-use. *Journal of Computer-Mediated Communication, 12*(2), 635-651.

Cummings, J., Finholt, T. A., Foster, I., Kesselman, C., & Lawrence, K. A. (2008). Beyond being there: A blueprint for advancing design, development and evaluation of virtual organizations. Arlington, VA: National Science Foundation.

Cummings, J., & Kiesler, S. (2003). KDI initiative: Multidisciplinary scientific collaborations (NSF Award No. II-S 9872996). *National Science Foundation.*

Cummings, J., & Kiesler, S. (2005). Coordination and success in multidisciplinary scientific collaborations. *Social Studies of Science, 34*(5), 1-20.

Cummings, J., & Kiesler, S. (2007). Coordination costs and project outcomes in multi-university collaborations. *Research Policy, 36*(10), 1620-1634.

David, P. A., & Spence, M. (2003). Towards institutional infrastructures for e-science: The scope of the challenge. *OII Research Report No. 2*. Retrieved April 6, 2014, from http://www.oii.ox.ac.uk/resources/publications/RR2.pdf

De La Flor, G., Jirotka, M., Luff, P., Pybus, J., & Kirkham, R. (2010). Transforming scholarly practice: Embedding technological interventions to support the collaborative analysis of ancient texts. *Computer Supported Cooperative Work, 19*(3), 309-334.

DiMaggio, P., & Powell, W. W. (Eds.) (1983). The *new institutionalism in organizational theory*. Chicago: University of Chicago Press.

Dourish, P., & Bellotti, V. (1992). Awareness and coordination in shared workspaces. Proceedings of the 1992 ACM conference on Computer-supported cooperative work. Toronto, Ontario, Canada, ACM: 107-114.

Dourish, P., & Edwards, W. K. (2000). A tale of two toolkits: Relating infrastructure and use in flexible CSCW toolkits. *Computer Supported Cooperative Work, 9*(1), 33-51.

Edwards, P. (1999). Global climate science, uncertainty and politics: Data-laden models, model-filtered data. *Science as Culture, 8*(4), 437-472.

Edwards, P. N. (2010). *A vast machine: Computer models, climate data, and the politics of global warming*. Cambridge, MA: MIT Press.

Epstein, S. (1996). Impure science: AIDS, activism and politics of knowledge. Berkeley, CA: University of California Press.

Fussell, S., Kraut, R., Siegel, J. (2000). Coordination of communication: Effects of shared visual context on collaborative work. *Proceedings of the 2000 ACM conference on Computer supported cooperative work, Philadelphia, Pennsylvania, United States*, 21-30.
Glaser, B. G. (1978). *Theoretical sensitivity: Advances in the methodology of grounded theory*. Mill Valley, CA: Sociology Press.
Glaser, B. G., & Strauss, A. (1973). *The discovery of grounded theory: Strategies for qualitative research.* Chicago: Aldine Pub. Co.
Goodwin, C. (1995). Seeing in depth. *Social Studies of Science, 25*(2), 237-274.
Graham, S., & Thrift, N. (2007). Out of order: Understanding repair and maintenance. *Theory, Culture & Society*, *24*(3), 1-25.
Hedstrom, M. (2001). Exploring the concept of temporal interoperability as a framework for digital preservation. *Proceedings of the Third DELOS Network of Excellence Workshop on Interoperability and Mediation in Heterogeneous Digital Libraries, Darmstadt, Germany.*
Hine, C. (2006). Databases as scientific instruments and their role in the ordering of scientific work. *Social Studies of Science, 36*(2), 269-298.
Hollan, J., & Stornetta, S. (1992). Beyond being there. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Monterey, California, USA,* 119-125.
Howison, J., & Herbsleb, J. D. (2011). Scientific software production: Incentives and collaboration. *Proceedings of the 2011 ACM Conference on Computer Supported Cooperative Work, Hanzhou, China*, 513-522.
Hughes, T. P. (1999). *Funding a revolution: Government support for computing research*. Washington, DC: National Academy Press.
Igo, S. E. (2007). *The averaged American: Surveys, citizens, and the making of a mass public*. Cambridge, MA: Harvard University Press.
Jackson, S. J., Pompe, A., & Krieshok, G. (2012). Repair worlds: Maintenance, repair, and ICT for development in rural Namibia. *Proceedings of the 2012 ACM Conference on Computer Supported Cooperative Work, Seattle, Washington, USA,* 107-116.
Jackson, S. J., Ribes, D., Buyuktur, A., & Bowker, G. C. (2011). Collaborative rhythm: temporal dissonance and alignment in collaborative scientific work. In *Proceedings of the ACM 2011 conference on Computer supported cooperative work* (pp. 245-254). ACM.
Jackson, S. J., Steinhardt, S. B., & Buyuktur, A. (2013). Why CSCW needs science policy (and vice versa). In *Proceedings of the 2013 conference on Computer supported cooperative work* (pp. 1113-1124). ACM.
Karasti, H., Baker, K. S., & Millerand, F. (2010). Infrastructure time: Long-term matters in collaborative development. *Computer Supported Cooperative Work*, *19*(3-4), 377-415.
Keating, P., & Cambrosio, A. (2003). *Biomedical platforms: Realigning the normal and the pathological in late-twentieth-century medicine*. Cambridge, MA: MIT Press.
Kee, K. F., & Browning, L. D. (2010). The dialectical tensions in the funding infrastructure of cyberinfrastructure. *Computer Supported Cooperative Work*, *19*(3-4), 283-308.
Kelle, U. (2007). The development of categories: Different approaches in grounded theory. In A. Bryant and K. Charmaz (Eds.), *The SAGE handbook of grounded theory* (pp. 191-213). Thousand Oaks, CA: Sage.
Kellogg, K. C., Orlikowski, W. J., & Yates, J. (2006). Life in the trading zone: Structuring coordination across boundaries in postbureaucratic organizations. *Organization Science, 17*(1), 22-44.
Knorr-Cetina, K. (1999). *Epistemic cultures: How the sciences make knowledge*. Cambridge, MA: Harvard University Press.
Kraut, R., Egido, C., & Galegher, J. (1988). Patterns of contact and communication in scientific research collaboration. *Proceedings of the 1998 ACM Conference on Computer Supported Cooperative Work, Portland, Oregon, United States*, 1-12.
Kuhn, T. S. (1962). The structure of scientific revolutions. Chicago: Chicago University Press.
Lamont, M. (2009). *How professors think: Inside the curious world of academic judgment*. Cambridge, MA: Harvard University Press.
Latour, B. (1987). *Science in action: How to follow scientists and engineers through society*. Cambridge, MA: Harvard University Press.

Latour, B. (1999). *Circulating reference: Sampling the soil in the Amazon forest*. In B. Latour (Ed.), *Pandora's hope: Essays on the reality of science studies* (pp. 24-79). Cambridge, MA: Harvard University Press.
Lawrence, K. A. (2006). Walking the tightrope: The balancing acts of a large e-research project. *Computer Supported Cooperative Work* ,*15*(4), 385-411.
Lawrence, T., Suddaby, R., & Leca, B. (2011). Institutional work: Refocusing institutional studies of organization. *Journal of Management Inquiry, 20*(1), 52-58.
Lee, C. P., Dourish, P., & Mark, G. (2006). The human infrastructure of cyberinfrastructure. In *Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work* (pp. 483-492). ACM.
Mackie, C. J. (2007). Cyberinfrastructure, institutions, and sustainability. *First Monday*, *12*(6), 1-8.
Millerand, F., Ribes, D., Baker, K. S., & Bowker, G. C. (2013). Making an Issue out of a standard: Storytelling practices in a scientific community. *Science, Technology & Human Values*, *38*(1), 7-43.
Monteiro, E. (1997). Scaling information infrastructure: The case of the next-generation IP in the Internet. *The Information Society, 14*(3), 229-245.
Nardi, B., Whittaker*, S.,* & Bradner, E. (2000). Interaction and outeraction: Instant messaging in action. *Proceedings of the 2000 ACM Conference on Computer Supported Cooperative Work, Philadelphia, Pennsylvania, United States*, 79-88.
Olson, G. M., Finholt, T. A., & Teasley, S. D. (2000). Behavioral aspects of collaboratories. In S. Koslow & M. F. Huerta, *Electronic collaboration in science* (pp. 1-14 ). Mahwah, New Jersey: Lawrence Erlbaum Asociates.
Olson, G. M., & Olson, J. S. (2000). Distance matters. *Human-Computer Interaction, 15*(2-3), 139-179.
Olson, G. M., Zimmerman, A., & Bos, N. (Eds.) (2008). *Scientific collaboration on the Internet*. Cambridge, MA: MIT Press.
Pollock, N., & Williams, R. (2008). *Software and organizations: The biography of the enterprise-wide system or how SAP conquered the world.* New York: Routledge.
Pollock, N., Williams, R., & D'Adderio, L. (2007). Global software and its provenance generification work in the production of organizational software packages. *Social Studies of Science*, *37*(2), 254-280.
Ribes, D. (2014). The Kernel of a Research Infrastructure. *Proceedings of the ACM 2014 Conference on Computer Supported Cooperative Work (CSCW), Baltimore, MD*, 574-587.
Ribes, D., & Bowker, G. C. (2008). Organizing for multidisciplinary collaboration: The case of the geosciences network. In G. M. Olson, A. Zimmerman, & N. Bos, *Scientific collaboration on the Internet* (pp. 311-330). Cambridge: MIT Press.
Ribes, D., & Bowker, G. C. (2009). Between meaning and machine: Learning to represent the knowledge of communities. *Information and Organization, 19*(4), 199-217.
Ribes, D., & Finholt, T. A. (2009). The long now of technology infrastructure: Articulating tensions in development. *Journal for the Association of Information Systems*, *10*(5), 375-398.
Ribes, D., Jackson, S. J., Geiger, S., Burton, M., & Finholt, T. (2012). Artifacts that organize: Delegation in the distributed organization. *Information and Organization, 23*(1), 1-14.
Ribes, D., & Polk, J. B. (2012). Historical ontology and infrastructure. *Proceedings of the 2012 iConference*, *Toronto, CA,* 252-264.
Star, S. L. (1999). The ethnography of infrastructure. *American Behavioral Scientist*, *43*(3), 377-391.
Star, S. L., & Griesemer, J. R. (1989). Institutional ecology, "translations" and boundary objects: Amateurs and professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39. *Social Studies of Science*, *19*(3), 387-420.
Vertesi, J., & Dourish, P. (2011). The values of data: Considering the context of production in data economies. *Proceedings of the ACM 2011 Conference on Computer Supported Cooperative Work, New York, NY*, 533-542.
Zimmerman, A. (2008). New knowledge from old data: The role of standards in the sharing and reuse of ecological data. *Science, Technology & Human Values, 33*(5), 631-652.
Zimmerman, A., & Finholt, T. A. (2007, November). Growing an infrastructure: The role of gateway organizations in cultivating new communities of users. In *Proceedings of the 2007 international ACM conference on Supporting group work* (pp. 239-248). ACM.

## About the Authors

**David RIBES** is Assistant Professor in the Communication, Culture and Technology (CCT) Program at Georgetown University. He is a sociologist of science who focuses on the development and sustainability of research infrastructures (i.e., networked information technologies for the support of interdisciplinary science), their relation to long-term changes in the conduct of science, and epistemic transformations in objects of research. He has a degree in sociology, but the field of science and technology studies (STS) is his first affiliation. His methods are ethnographic, archival, and comparative. See davidribes.com for more.

**Jessica Beth POLK** is a doctoral student in the Department of Science & Technology Studies (S&TS) at Cornell University. Her research stands at the intersection of the sociology of science (with a focus on medicine), infrastructure studies and systems of data sharing, as well as the study of medical expertise and experience. Prior to obtaining her MA at Georgetown University's Communication, Culture & Technology Program (CCT), she spent several years working in the healthcare industry.