

STS, Meet Data Science, Once Again

Science, Technology, & Human Values

2019, Vol. 44(3) 514-539

© The Author(s) 2018



Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/0162243918798899

journals.sagepub.com/home/sth**David Ribes¹****Abstract**

Science and technology studies (STS) and the emerging field of data science share surprising elective affinities. At the growing intersections of these fields, there will be many opportunities and not a few thorny difficulties for STS scholars. First, I discuss how both fields frame the rollout of data science as a simultaneously social and technical endeavor, even if in distinct ways and for diverging purposes. Second, I discuss the *logic of domains* in contemporary computer, information, and data science circles. While STS is often agnostic about the borders between the sciences or with industry and state—occasionally taking those boundaries as an object of study—data science takes those boundaries as its target to overcome. These two elective affinities present analytic and practical challenges for STS but also opportunities for engagement. Overall, in addition to these typifications, I urge STS scholars to strategically position themselves to investigate and contribute to the breadth of transformations that seek to touch virtually every science and newly bind spheres of academy, industry, and state.

¹Human Centered Design and Engineering (HCDE), University of Washington, Seattle, WA, USA

Corresponding Author:

David Ribes, Human Centered Design and Engineering (HCDE), University of Washington, 428 Sieg Hall, 3960 Benton Lane NE, Seattle, WA 98195, USA.

Email: dribes@uw.edu

Keywords

data science, universal science, domain, STS, social science, cyberinfrastructure

Introduction

In recent meetings, conferences, and workshops, I have been taken by how many science and technology studies (STS) scholars are discussing data science (under its various namings: big data, data intensive discovery, machine learning, etc.) and far more strikingly how many of those scholars are not only treating it as an object of investigation but are in some way entangled with its rollout, for example, shaping curricula, participating in projects, collaborating in design, or in evaluative and policy roles. In part, this is because data science, at least in some spheres, is cast not only as a narrow matter of technical advance but also as a pedagogical, collaborative, and organizational endeavor, and, occasionally, as a reflexive conversation about epistemological transformations, the social good, ethical responsibility, and downstream consequences. In many ways, in these various niches, STS scholars (broadly construed) have engaged data science. Perhaps data science is simply a new bandwagon for STS, an impactful banner or just a hot and trending term that is receiving rapid research uptake, but here I will argue otherwise. STS and data science have elective affinities that are long coming, will offer many interesting cross-cutting opportunities, and pose some thorny challenges for STS (and certainly for data science, but I will leave those for another time).

There is a great deal of discussion and debate today about what is data science or who is a data scientist. I will explore several characteristics in this essay, particularly focusing on the keyword “domain” and its associated logic, but as a starting point, the recent definition from the Computing Research Association (CRA 2016) serves as well as any other: “we use the term data science in its broadest sense, including data collection, data engineering, data analytics and data architecture.” The general tone of the field is to take “data” as the object of study, a new *science of the artificial* (Simon 1980) that will investigate data’s regularities *as data* and, notably, with a strong engineering mindset to facilitate their transformation, interoperation, and representation. Data science is experiencing a gold rush. We are witnessing an Abbott (1988) style disciplinary turf war, with computer science, statistics, mathematics, information science, and various collaborations with scientific, state, and industrial sectors each making claims to the

expertise, funding, and regulatory sphere of data science. No single configuration seems to hold, and at each academic institution I have visited or inspected, I have found a different arrangement of interlinked disciplines and sectors (and those left out or cut out) of the game.

The outcome of the turf war, however, has a great deal more at stake than a slice of the disciplinary pie. Data science is positioning itself as a *universal science*, essential to all future sciences and beyond, connecting to and informed by industry and state. As Bowker (1993) once wrote of cybernetics, it seems data science could be substituted for that word at each instance:

cybernetics could operate either as the primary discipline, directing others on their search for truth, or as a discipline providing analytic tools indispensable to the development and progress of others. At both the superstructural and infrastructural level, the rhetoric held that cybernetics was unavoidable if one wanted to do meaningful, efficient science. (p. 22)

The wave of data science's institutionalization across the academy, particularly through new disciplinary programs, centers or certificates, degrees, and changing curricula, indicates a long-term investment in research and training in data science. No doubt the term is receiving such institutional uptake because of national agency, private foundation, and business-based backing that are emerging to and fro, and the double-impetus that industry is lending the term by generating a "push" to train new students and a "pull" through the promise of new jobs. None of this makes the rise of data science as the key header any less real; it is precisely one facet of how it is becoming more so.

Even if the term "data science" does not stick, the general trends I describe here precede that naming and are reflective of a deeper logic and trajectory (Leonelli, Rappert, and Davies 2016; Pollock and Williams 2008). Below I address two elective affinities: the *always already social of data science* and the *logic of domains*. In this essay, I will trace back only the thread most familiar to me through fieldwork in the United States during the 2000s and 2010s: the policy paradigm called cyberinfrastructure (CI), which shares some goals, actors, and many institutionalization strategies with data science. Within CI, collaborations with social scientists became relatively common, often in "embedded" relationships. These relationships seem to be even more common, and diverse, in contemporary data science. Data science is not a solely academic enterprise; in fact, a common refrain is that it is rooted and stronger in industry. However, in this essay, I focus almost entirely on academic manifestations, the settings I understand best.

Tracing this lineage reveals a fragment of the long converging trajectory of digital STS scholars and computing practitioners, in a lengthy lineage of computing intensive endeavors. I trace this historical lineage from CI to data science in order to draw attention to their continuities—in particular, the envisioned utility of social science—but also their discontinuities—in particular, a magnified attention to the social consequences of technological innovation. In the conclusion, I note how these elective affinities open opportunities for engaged scholarship and shaping the sociotechnical roll-out of data science, and argue that the opportunities for the field of STS (writ large) are growing in this milieu.

While there is a burgeoning research program on the public consequences of data (Gillespie 2010; boyd and Crawford 2012)—such as social media, financial, or policing data—the ambitions of data science are broader and deeper, often operating distally to publicly proximate data and their consequences. The methods, concepts, and skills of STS researchers place the field in a superb position to inspect these deep technicalities and logics as they are developed far from the identifiable consequences of data, but that thereafter travel freely on the increasingly well-paved roads of interoperability (Ribes 2017).

I'll conclude with some notes about the opportunities and forthcoming challenges for STS as engaged participants and scholars in this emerging space. The two elective affinities between data science and STS that I recount here offer avenues for STS scholars to participate and shape data science as a matter of technical design, policy consideration, or pedagogical planning. These opportunities also present several dangers; here, I will focus on the risk of reinscribing raw distinctions between the social, the technical and scientific, and the structured propensity for social science (and thus STS) to be positioned as a “service science.”

Our Elective Affinities

An elective affinity describes a relationship in which sets of actors who share certain analogies and meanings (religious, intellectual, or, in this case, sociotechnical) enter into a relationship of reciprocal attraction and influence, mutual selection, active convergence, and mutual reinforcement. “Elective affinities” enter the social theoretical discourse through the work of sociologist Max Weber.¹ Most notably, he argued that certain work and fiduciary ethics in Protestantism had promoted the primitive accumulation of wealth needed to jump-start capitalism (Weber [1904] 2013). Gerth and Mills (1946) describe elective affinities as “the decisive

conception by which Weber relates ideas and interests” (p. 62). That is, Protestantism and capitalism did not initially need to share the same logics, interests, or even actors in order to mutually buttress each other. Similarly, while STS (writ large) and data science (also writ large) approach science, expertise, and practice via differing routes and for distinct purposes, they share elective affinities that have mutually supported each other for decades and increasingly.

The roots of the elective affinities between STS and data science are multiple. The role of various stripes of social scientists and humanists has been well-documented in many threads of computing such as the participation of anthropologists and psychoanalysts in the Macy Group so influential to cybernetics (Heims 1991), the early formation of successful software engineering as a matter of social organization (Slayton 2013), or how branches of my own field of sociology such as symbolic interactionism and ethnomethodology have played direct roles in the design of interactive systems, human–computer interaction, or embodied and ubiquitous computing (Vertesi et al. 2016; Dourish 2001). It is not the purpose of this essay to root out all these rhizomes; suffice it to say that in many ways, humanists, social, and computing scientists have long been intersecting, collaborating, critiquing, or informing each other.

While in this essay I trace two elective affinities between STS and data science, I am not yet prepared to make a strong claim about the nature or mechanisms of those relationships. This paper is *programmatic* in the sense that I expect to continue the investigations outlined here; in this sense, this essay posits objects of investigation rather than a definitive theorization about them, in the hope of opening avenues of investigation and engagement for others. Calling something an elective affinity helps identify and characterize a phenomenon, without necessarily offering an explanation. I will, however, point to some of the ways that data science (and its antecedents) has defined a utility in, on the one hand, social scientific theorizing about the science, technology, and society relationship and, on the other hand, empirical work that characterizes particular expert communities (or “domains”). In turn, I also point to some of the specific intersections where computational research on the nature of knowledge, science, expertise, ontology, epistemology, and domains have at times intrigued, even inspired, STS scholars, thus furthering the elective affinities. The arguments here are based on investigations focused on American academic projects and science policy over the last twenty years, and they anticipate a broader historical, cross-sectoral, and international analysis.

The Always Already Social of Data Science

I was recently approached by a National Science Foundation (NSF) program officer to conduct a study of the Big Data Regional Innovation Hubs and Spokes (fondly known as BDHubs), a new national umbrella for American data science. That officer sought to entice me in an e-mail: “I think this is a unique opportunity to see how such research collaboration infrastructures evolve. The four Hubs have been funded for the same purpose, but each has its own character and internal organization, etc.” I have studied many similar endeavors in the past, but this is the first time that I was approached to do so—with an avenue for funding and generous access. In the past, I have approached them. However, being approached is not an altogether uncommon occurrence in my subfield; a handful of my colleagues have experienced something similar. Just over a decade ago, my then doctoral advisor, Geoffrey Bowker, had been similarly approached to do a study of GEON—“Cyberinfrastructure for the earth sciences.” That project became my dissertation.

It was the very same person who had approached Geoffrey Bowker who recently once again approached both of us to study BDHubs. He noted of BDHubs, “This reminds me a bit of the GEON early days . . . but now we truly have this at a national scale.” Once a “cyberinfrastructure person” at the San Diego Supercomputer Center, he has now become a “data science person” at the NSF; once a technology developer, he is now closer to a science policy actor who has sustained a common vision for scientific computing across these roles. In tandem, whereas at one point I identified as doing CI studies, a decade later, I suppose I am now doing data science studies.

I’ll return to these traveling companions later in this essay. Here, I discuss why a qualitative and openly epistemologically and ontologically concerned sociologist and STS scholar such as myself once offered some appeal to CI and now to data science practitioners. In short, one answer is that “the social” was an explicit matter of concern in CI and today far more so in many data science circles.

It is worth taking a brief moment, without delving too far into theory, to describe my approach to “the social” in the essay. I am *not* taking the social to refer to some ever-present, always underlying, fundamentally consistent, and unchanging feature of the interactional human world. Instead, broadly inspired by Foucault ([1975] 1995) and Rose (1990) and more proximate to STS, by Knorr-Cetina (1997), I am here treating *the nature of the social* as the outcome of particular historical–institutional arrangements. Framed

broadly, the social is the product of boundary work (Gieryn 1983; Latour [1991] 1993) that distinguishes, in this case, the social from the technical and sometimes from the scientific. That boundary is arguably one central object of STS scholarship, which in various ways scholars have at times sought to shift or even dissolve, but that in all cases can be considered the outcome of negotiation and contestation. More specific to this essay, I am interested in how various policy paradigms (Epstein 2007)—specifically CI and data science—have defined the utility of the social, and at times of social scientists, for successfully deploying novel technology and for managing the challenges of collaboration, interdisciplinary, or large-scale organization as matters of “social” concern.

I consider CI as an era of American science and technology policy, roughly beginning with the publication of the NSF report *Revolutionizing Science and Engineering through Cyberinfrastructure* in 2003 (better known as the Atkins [2003] Report).² What followed, in the United States (and through sister endeavors in Europe), was the deployment of multiple infrastructure development projects, primarily but not exclusively for the natural and engineering sciences. GEON, the ethnographic object of my dissertation, was one such CI, and in that project, I was one such “social scientist.”

In the CI policy paradigm, the nature and utility of the social and social scientists were articulated with a particular vision, largely cast as a matter of understanding and facilitating human collaboration and coordination. For instance, the Atkins (2003) Report identified as key troubles for the envisioned rollout of CI:

lack of appreciation of social/cultural barriers, lack of appropriate organizational structures, inadequate related educational activities, and increased technological (not invented here) balkanizations rather than interoperability among multiple disciplines. (p. 4)

Common formulations of the role of social science included investigating and helping with, for example, working in a geographically distributed manner, bringing together heterogeneous disciplinary scientists, or examining the challenges of data sharing. The framing of these social–organizational questions was largely internal to the goals of CI, that is, linked to the question of successfully developing national, interdisciplinary, computational networks to support science. This framing was coupled with the participation of a heterogeneous handful of “social scientists,” which

became interlinked with the rollout of various CI era projects; sometimes studying them, at others assisting them, often both.

The participation of social scientists was “built into” the model of CI. For instance, with the Atkins (2003) Report reminding its readers of “social science research communities who see [Cyberinfrastructure] as an *object of research*” (emphasis in original) and arguing that developing CI “effectively requires holistic attention to mission, organization, processes, and technology. It creates the need to involve social scientists as well as natural scientists and technologists in a joint quest for better ways to conduct research” (p. 26). I return to this tripartite distinction—scientists, technologists, and social scientists—in the following sections.

More proximate to the genealogies of STS, it is no coincidence that Geoffrey Bowker and I were approached to do studies of GEON and BDHubs: just over two decades before, Star and Ruhleder’s (1994) investigation of the WORM Community infrastructure had instituted what has now become a thriving research program and redefinition of infrastructure (i.e., as a relation, not a thing). Star had also made a contribution to studies of collaboration, for instance, as requiring neither conceptual nor interest-based consensus, an argument summed up in the well-known concept of the “boundary object”—more on this later. Many others at the intersection of information studies and STS have contributed to a framing of infrastructure as a matter of organization, communication, and collaboration (Edwards et al. 2013; Jackson et al. 2007). In various ways, these concepts and findings have been taken up by technologists *and* STS scholars.

Turning now to data science, I find that the pragmatic or “functional” (Berg 1998) questions of collaboration, coordination, and social organization that infused CI’s social research remain, more or less, going concerns in contemporary data science endeavors. As Szalay and Feldman (2018) recently put it: “The challenge is more in social engineering than a technical one,” a recurrent phrase I have heard since beginning my fieldwork with GEON. However, a far more expansive social agenda has emerged: within data science circles, the role of the social has been significantly magnified and diversified. I will illustrate this with a brief examination of the data science institute at my own institution, the University of Washington:

The eScience Institute, with the tag-line “advancing data-intensive discovery *in all fields*” (emphasis added), has a formal organizational structure of five workgroups. Of those five, only one could be considered narrowly technical, the other four have pedagogical, institutional, or reflexive roles: Software Tools, Environment, and Support; Career Paths and Alternative

Metrics; Education and Training Workgroup; Reproducibility and Open Science; and Data Science Studies.

Perhaps the *Software Tools, Environment and Support* workgroup is the most expected, targeting the development of data analytic tools for and across specific domains. But there is also the *Career Paths and Alternative Metrics* workgroup that tackles the challenges of finding academic homes for highly interdisciplinary data scientists within a hiring ecology defined by disciplines. That group also seeks to foster transverse relationships between the academy and industry. There is also the *Education and Training* workgroup that regularly organizes events such as *hackathons* or *Software Carpentry* lessons that seek to entrain scientists at any level of computational expertise in novel data analytic skills. These workgroups have missions that combine a technical set of goals with programmatic visions, such as the *Reproducibility and Open Science* workgroup that develops tools (and an ethos) that seek to facilitate the easy and accountable flow of data. These latter three working groups have only some technical ambitions; they can more tidily be described as furthering the institutional, even normative, goals of data science.

The *Data Science Studies* workgroup is perhaps most intriguing to the readers of this journal, defining itself as “a group of cross-disciplinary researchers studying the sociocultural and organizational processes around the emerging practice of data science.” That working group is led by a *data science ethnographer* and a *human-centered data scientist*, and the eScience center has supported several graduate student ethnographers-in-training. The group regularly organizes events that bring together STS and data science researchers along with state and industry actors, to reflect on historical, present, and future challenges of data science. I will give a further taste of what these ethnographers, including myself, do in data science projects, but that list is expanding daily, and I am certain that no accounting today will capture our entanglement tomorrow.

Beyond organizational or pragmatic considerations, the contemporary concern with data and their consequences has vastly broadened what was once CI’s model of the social with a much wider agenda. In short, and in a way that is quite different from CI, today *data science has encountered Society*. While CI framed itself as largely targeting the sciences, data science sees no such boundaries. Data science is a term that was taken up much faster in Silicon Valley than the academy but today also in state agencies and nongovernmental organizations. Herein lies the expansion of the social for data science. In the last decade, we have observed both a revived public, scholarly, and activist recognition that data matter and the

proliferation of instruments (and thus data) within virtually every aspect of private and public life, globally even if unevenly. Data, their circulations, transformations, and interoperations, have been “politicized”—in both the literal and figurative senses of the word—and that politicization has been coupled with the rise of a new set of activist, watchdog, regulatory, scholarly, and political organizations, along with a reorientation of multiple long-standing civil society groups concerned with the press, civil liberties, privacy, surveillance, and so on.

Today, only a data scientist playing the ostrich will fail to recognize that the flows and combinations of data are societally consequential. The result has been a take-up of the concerns and language of social justice (among others) as part of some data science ventures. Technical formulations of social problems associated with data—that is, privacy or security—are part of all data science ventures, and many go further by explicitly calling out classical social justice questions of resource distribution, inclusion, and bias.

Summing up, I am arguing that while explicit formulations of the social were not absent in CI, we are witnessing a significant expansion of its enactment in data science: a coupling of organizational, institutional, pedagogical, design, and ethical concerns tied together in a programmatic vision of technological rollout. In such niches, STS scholars have found many a role, that is, participatory, leaderly, critical, evaluative, and so on. The formulation of the social in STS and in data science do not fully cohere, but as we will see, there are elective affinities in the terms, meanings, and methods of inquiry that serve to sustain and renew opportunities for engagement.

To be clear, I am not making an evaluative judgment in pointing to the various political, social, or social scientific framings in data science. I am less concerned in this essay with whether the endeavor is correct or sufficient in its understanding of the social. Instead, I am noting a particular institutional configuration that has generated a kind of concern for the social, and the utility of social science, primarily as a matter of successful technoscientific development and, more recently, as a concern for broader societal consequences.

What is a Domain? Boundary Work and its Crossing

We could call it a self-similar construct, a figure whose organising power is not affected by scale.

Strathern (1995, 18)

For many years in my writing, I have used an *in vivo* category “domain” and especially “domain scientist”—terms drawn from my fieldwork. I am in the process of assembling a longer genealogy for the term, but by the time I began my fieldwork in 2003, “domain” was in common usage among scientists collaborating to develop CI. This actors’ formulation precedes CI—it is more broadly used within the computer and information sciences, as well as business, librarianship, and other fields—but the implicit vision appears to be reaching a new fruition with data science. Domain is a keyword for data science (Williams 2014 [1976]) and the associated fields that preceded it. The distinction between that which is or is not a domain serves as an organizing principle for a universal science. And, since some part of what characterizes a domain is within the purview of social science, the logic of domains articulates a role or utility for social science.

Today the usage of domain is often colloquial rather than formal. While there is an extensive technical literature on domains and their analysis (cf. Hjørland and Albrechtsen 1995; Neighbors 1980), the term domain itself has transcended these discussions and become a vernacular expression in data science circles. It is used in everyday talk, but also in policy or scientific publications, displaying significant malleability. When I explicitly probe the meaning of domain, it evokes debate not about the term itself but about the role of domains in data science or computation more generally. Here, I paint an impressionist image of that term and in the next section articulate the particular utility that the logic of domains finds in the work of social scientists.

Very roughly, domains refer to those fields (often scientific, but not exclusively) concerned with worldly and specific matters, for example, linguistics is the “domain science” of language, biologists are the “domain experts” of organic life, and so on. The logic of domains parses the world into two main categories, one is either “in a domain” or one is working “independently” of any domain. During my fieldwork in the 2000s, the category “domain scientist” was quite definitive: in my study of GEON, saying domain scientist was equivalent to saying geoscientist, and everyone else was a computer scientist or technologist (except me—more on this below). Today the distinction is not nearly so stark. Many domain scientists *identify as* being data scientists: a data scientist may *be* a geoscientist. But the term domain still does work, perhaps more work, by defining or constituting a terrain between or beyond individual domains as the ultimate target for data science.

There are also nondomain concerns, modes of expertise, and technologies. But that position is unnamed and unmarked, often negatively defined

as, for instance, “domain independent.” For example, a domain-independent data analysis tool might find utility within geoscience *and* biology. Who or what specialization is doing domain-independent activities is a matter of debate but often includes some composition of statistics, information science, mathematics, and today it seems to invariably include computer science. Could one say that data, computer, or information sciences too are domains, with their own expertises and practices? One could, but it is rarely how the expression is used.³ In parlance, data, computing, and information sciences are “the other” to the domains.

In this sense, data science is in the intellectual, practical, and disciplinary lineage of statistics, network analysis, computer and information science, and perhaps most strongly, cybernetics—and in ways comparable to these fields, their objects of study can be found anywhere (respectively: ordered numbers, networks, algorithms and information, feedback loops, and now data). In a paper that catalogued its tricks and strategies, Bowker (1993) cast cybernetics as a field that sought to position itself as a universal science:

Cybernetics, through its universal language, described what could in the broadest sense be called “a new economy of the sciences”[. . .] it sought to order the sciences in a different manner [. . .] by simultaneously offering new ways in which they could cognitively interact with each other, and establishing new sources of funding to facilitate these interactions. In this new economy [. . .] chemists were empowered to do things that normally only engineers could do. And they could do it because they could tap into biological ideas through the mediation of cybernetics. Anyone tapping into the network of words used by cybernetics would be tapping into the network of problems that cyberneticians were aiming to solve. (p. 117)

Bowker noted several interleaving thrusts of the cybernetics program: a universal language that operates across (or above) disciplinary languages, tools and techniques that do the same, and an economy or ecology that fosters new arrangements across the disciplinary divides of the sciences. The term domain serves a similar role: they are to be found everywhere and anywhere that expertise can be characterized as distinct, whether those differences are practical, conceptual, or epistemic.

Importantly, at its starting point, the logic of domains has nothing specific to say about a particular domain or how it may differ from others. That question is fully empirical, to be known via the task of investigating documents, data, or by collaborating with the members of that domain. As

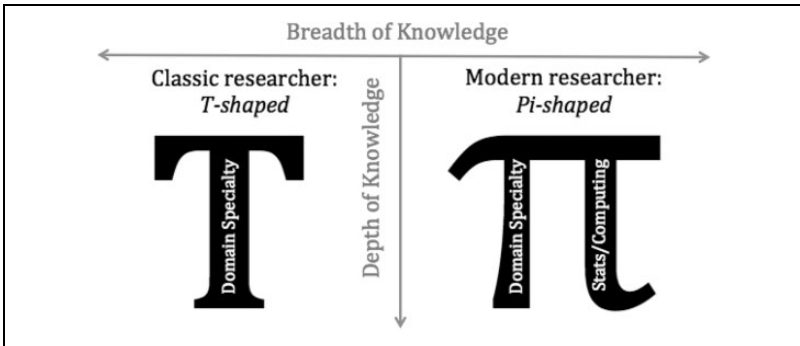


Figure 1. The T and “ π -shaped researcher.” This meme-like formulation circulates widely within data science discussions.

Bowker (1993) noted, “Both dimensions (content-rich and content-free) were integral to the success of the new universal language” (p. 117). Domain is just such a term, packed with importance but also nothing specific. The distinction domain/independent is a key organizing principle for the rollout and practice of data science and simultaneously marks an object of inquiry, of service, and of intervention, a distinction between the specific and the general, and sometimes the specific and the universal.

The space between domains and nondomains is not unpopulated; in fact, it is often lionized. A current meme-like formulation in data science is the “ π -shaped person” (see Figure 1) with each leg of π representing deep knowledge and the squiggle the expertise to bring them all and bind them. So, a π -shaped person is an individual with expertise in both data science techniques and a domain expertise (say, a data intensive geoscientist). The phrase is usually attributed to astronomer and computer/data scientist Alex Szalay, though several other letter-shaped people in private industry spaces preceded it. The explicit contrast is with a “T-shaped person” with a single specialization and a breadth of general understanding. Sometimes the triple pronged “m-shaped person” is deployed as an even greater ideal, adding an additional domain or programming and hacking expertise to the other specializations. The various kinds of symbol-shaped people continue to proliferate and have become a vernacular trope and concept-to-think-with in data science circles. All these formulations rely on some form of the domain/independent distinction.

One goal of data science is to establish transverse relationships in the sciences as well as with state and industry (all of these are regularly cast as

domains). In academic settings, these are sometimes described as disciplinary boundaries; in private industry as sectors; for data and software they are stovepipes or silos, terms that normatively hint at the challenges of collaborating or sharing across such boundaries. Data science does not seek to disassemble the domains but rather to facilitate and create tools or data-based relationships across them. Ultimately, the object of data science is not a domain; the term serves to identify that which will be *served and transformed* by the tools of data science or crossed with another domain, several of them, or all of them. Ultimately, goals for domain independence seek to transcend the specificity of domains.

While I cannot yet fully articulate the logic and practice of domains, I can offer three analytic starting points: domains are the outcome of boundary work but with the goal of crossing those boundaries, the logic of domains is distinct from the logic of multidisciplinary, and a remarkable feature of domains is that they are *holographic or fractal*.

First is to consider domains as the outcomes of boundary work (Gieryn 1983) but with the purpose of facilitating the crossing of those boundaries. As Gieryn argued, boundary drawing is practical and rhetorical work and can be taken as standing in for an interest to define expertise, secure resources, or protect jurisdictions. However, more important than defining an inside and an outside, boundary work in the logic of domains serves to characterize difference for the purpose of overcoming boundaries. Whether it be technical, cultural, or practical boundaries, these are all articulated for the purpose of finding ways to create bridges. Domains are framed as distinct, not (only) for the purpose of securing resources or prestige, as with the initial formulations of boundary work, but in order to identify challenges and promising pathways for crossing to another domain. Domains are one of the key empirical “working objects” (Daston 2000) for the sciences and engineering of computation, information, and data.

A second consideration is that the logic of domains is distinct from what we might call the logic of multidisciplinary, even while often operating in tandem with it. A characteristic feature of the logic of domains is positing (and then engineering) an intermediating object across discrete domains. Domains are thus not placed directly in contact, and their boundaries are not dissolved. Rather, difference is translated while leaving domains relatively unchanged. For instance, GEON sought to develop “computational ontologies,” mapping in formal representations the entities of each of its participating domains: for example, paleobotany, geophysics, and seismology. These diverse scientists did not come together to hammer out a single common ontology; they each developed their own and helped to craft

automated translations across them (Ribes and Bowker 2009). The ultimate goal was to develop pathways by which, say, a geophysicist could draw on the data of seismology, without having to learn the specific (unfamiliar and esoteric) categories by which geophysical data are generated and organized—a form of data interoperability (Ribes 2017). In contrast, multidisciplinary (or its cousins: inter- or transdisciplinarity, etc. [Klein 1996]) is usually characterized as a matter of collaboration in which different parties learn each other, for instance, sharing languages, methods, objects, or research questions. Such collaborations involve some form of direct interaction, exchange, and often learning by all parties. Multidisciplinarity seeks to soften or dissolve boundaries; the logic of domains seeks to preserve difference while connecting across it.

A final observation I can make is that domains have a holographic or fractal quality: they are found everywhere, even within themselves, and at any scale. For instance, GEON was “cyberinfrastructure for the geosciences,” but the term geoscience spanned broader than any university disciplinary department, with GEON tackling fields as diverse as geology, seismology, or paleobotany. In parlance, each of these was referred to as a domain as well and, as such, demonstrated undesirable difficulties in collaborating or sharing data with one another. The term domain can refer to a broad swath, top-level category, and its nested constituents.

Paraphrasing Strathern (1995) in her book *The Relation*, domain is “a self-similar construct, a figure whose organizing power is not affected by scale” (p. 18). In her writing, Strathern was theorizing the concept of “the relation” in general, but she was more specifically focusing on the “kinship relation” as an historical keyword for social anthropology. I see the role of domain in data science as a corollary to the role of kinship in social anthropology: they are concepts capable of mapping specific difference while doing so in the same way. Strathern noted that the social anthropologist could travel to any location and find human relations of kin: mothers and daughters, cousins and stepfathers. Kinship could be tailored to a particular culture, cutting some relations and adding others (i.e., add godfather, remove adopted child). Kinship is “fractal” in that it reprises at all scales, that is, kinship is a matter of lived practice but also encoded into law, taxation, or literature.

Domains operate in much the same way. They can be found at any scale, for example, the umbrella term “biology” can be cast as a domain but so can the much more specific field of “genomic epidemiology.” Domains are nested: by inspecting a domain further, more granular characterizations can be made of additional domains, posing additional challenges for working

together and additional grist for the work of data science. They are holographic, in that if you cut a hologram each part will contain the whole; similarly, if one analytically slices a domain, it will reveal further domains, each complete with its properties of having epistemologies, cultures, data, funding arrangements, and so on. Characterizing a domain can be a “technical” matter of understanding the domain’s data or software but also its more evanescent “social” features such as languages, cultures, or ontologies. For inquiring upon these, social scientists and their tools of inquiry have decades-long history of playing a role in the logic of domains.

The Role of the Social (Scientist) in the Logic of Domains

Throughout my projects in the 2000s, in the various meetings, conference calls, and hotel bars where I studied/collaborated with CI developers as a graduate student, they often referred to me as a “social scientist” (sometimes as “our social scientist,” more on this later). At first, I chafed at the term. As an STS-identified sociologist in training, the umbrella term social scientist was too broad for me to swallow; I recoiled at being collated with economists, social network analysts, and survey researchers (as any good disciplinarian would). Over time, though, I came to understand the work the term did in context. Social scientist served as a complement to the other two possible categories of people within CI efforts: domain scientists and computer scientists. As we have seen, both are vast umbrella terms referring, on the one hand, to any worldly sphere and, on the other, to the differing expertise of those who work on hardware, visualization, knowledge representations, and so on. In this triad, computer and information scientists were the developers of novel computational resources, domain scientists were those topical specialists who would be served by these new resources, and social scientists were those who took this sociotechnical development process as their object of study and, at times, the object of service. One such key investigative task is researching and characterizing the domains.

This role, or third position, remains a lively one today in data science circles. For instance, during the writing of this article, I received a request to fill in a cross-institutional survey of data scientists. I have written about the role of surveys in technology development projects, and how often (though not exclusively) it is “social scientists” who design and deploy them (Ribes 2014; Ribes and Baker 2007; Ribes and Finholt 2008). In large-scale technology development projects surveys are a key inquiry tool for understanding social, organization, and technical configurations of the prospective

user community (or domain). Thereafter, survey findings serve some role in informing downstream design of the computational system.

In this instance, the request came from a former student, Stuart Geiger, now himself a postdoctoral researcher at a data science institute. I was surprised to be one of the recipients (i.e., a subject) of this request to fill out a survey about data scientists and asked Stuart about his sampling method. He wrote to me:

We asked various people for lists of who was affiliated with or “in the orbit of” the [data science] institutes, and as I was cleaning and consolidating them, there was a moment when I saw STS/ethnography people and began taking them out of the sample. Then I thought: no, no, people we all recognize as Members named them as Affiliates, keep them in! (E-mail correspondence, October 10, 2016)

Notably, the “STS/ethnography people” inhabit a nether region of membership in data science; conducting a survey now requires this kind of boundary consideration in its design and rollout. As with survey results more generally in these information technology (IT) spaces, the downstream data and findings generated by this instrument will play some role in representing the diverging or aligning interests of domain, data science, and now STS/ethnography people within the milieu of data science. In the logic of domains, inquiry on the domains precedes intervention.

Beyond the empirical studies of domains, the theoretical innovations of STS have also gained traction within data and computational circles, particularly with respect to mediating domain relationships. I have found many STS originating concepts deployed within data science circles, such as pidgins and trading zones (Galison 1999), or human infrastructure (Lee, Dourish, and Mark 2006), but none so commonly as the boundary object, a notion which recapitulates the logic of domains by conceptualizing the preservation of difference through intermediation. It is worth taking a moment to reflect on how boundary objects have served as a conceptual resource in the logic of domains and, more importantly, how they have served as a *design resource in developing information systems*.

Hailing from the social worlds and arenas tradition of symbolic interactionism that matured within STS, the boundary object concept offered a distinct feminist critique of actor–network theory that sought to consider multiple parties (“an n-way translation”) rather than only victorious translation by managerial actors (“a one-way translation”). However, boundary objects were as much in discourse with computational thinking as with

feminism or actor–network theory: as Star (2010) wrote, “I use the term object in both its computer science and pragmatist senses” (p. 603). Notably, one of the two initial articles Star wrote on boundary objects in 1989 was published in the journal *Distributed Artificial Intelligence* and provided an alternate formulation of domain independence to that of cognitive scientist and artificial intelligence researcher Herbert Simon.

From this perspective, it should only come as a small surprise that data and information scientists have found boundary objects to be eminently and ideally operationalizable within their computational systems, for the concept asserts that heterogeneous communities can *collaborate without consensus* if a mediating artifact meets the informational requirements of all parties. This is a distinct model of the logic of domains. More than just a model, the boundary object concept offers an *ideal* for the logic of domains, that is, collaborating domain communities interfacing at an “object” may largely continue on their own trajectories rather than wholly transformed or integrated (n-way translation). If commercial trappers and scientific taxonomists can continue trapping and categorizing but in a coordinated fashion, then why can’t physicists and sociologists do the same?

Boundary objects have explicitly informed the design of many computational systems supporting collaboration or interoperability. For well-known examples, see Kjeldskov and Paay (2005); Simone, Mark, and Giubbilei (1999); Henderson (1991); Bergman, Lyytinen, and Mark (2007). Notably, these papers are not “studies of” or “theories of” boundary objects but rather written by system developers who cite the concept as a driving influence in the design their technical artifacts or tools. More generally, concepts originating from, or often appearing in, STS circles circulate with ease in the domain discourse of data science: for example, paradigms, pidgins, communities of practice, epistemic cultures, and human infrastructure. I have seen all of these used to explain domain boundaries, and deployed, in one fashion or another to ameliorate their crossings. These concepts appear to meet the informational needs of STS *and* the logic of domains, for differing purposes, but with elective affinities.

Whether it be Lucy Suchman or Julian Orr’s work at Xerox PARC, or Susan Leigh Star’s consultant-like role in developing nursing information systems and studying the WORM community, STS scholars have long engaged IT in ways other than simply “studying it” (Bowker, Timmermans, and Star 1997; Suchman 1987; Orr 1996; Vertesi et al. 2016). Beyond direct participation of STS scholars or the take-up of STS concepts in such design activities, methodologies such as ethnography (in its various forms of naming and disparate lineages), surveys, and other social science techniques

have decades-long histories of being incorporated into the discourse, curriculum, and practice of what has been variously called human factors, user/human-centered design, or user experience. Either through the active participation of social scientists or the take-up of their methods and concepts, the study of users, humans, contexts, or sociotechnical systems has proven a boon to social scientific method and inquiry, while also posing many a challenge. Here, I focus on one challenge and use it to stand-in for the larger complexities of collaborating in the logic of domains.

In a programmatic paper to the Human–Computer Interaction community, design scholars Paul Dourish and Graham Button (1998) noted that ethnomethodology has often been treated as a “service science, in which ethnomethodologists uncover requirements for system design or computer scientists develop systems organised around specific working practices” (p. 26). That paper, and many others (Berg 1998), has called for a broadening of the understanding and role for social science beyond “implications for design” (Dourish 2006). Cast widely: at any point that social science meets novel technology application demands, the possibility of playing a service role reemerges.⁴ The request for some contribution in exchange for field site access is an opportunity, and a challenge, for STS scholars. It is a challenge because contributing is hard, time-consuming, not in the immediate bailiwick of STS, and threatens to tip over into a service role. But at this intersection, we also find the chance to shape emergent sociotechnical landscapes, and this presents an opportunity for STS and allied fields. The two elective affinities I am describing here run deeper than individual engagements of STS scholars with data scientists (or their forerunners)—ties cannot simply be broken in some smashing of the idols. STS (writ large) in these spaces, and perhaps more generally, is conceptually, practically, and institutionally entangled in ways that I have only begun to unpack here, but, put more fruitfully, that entanglement is a first point of entry for shaping ongoing consequences.

How Should STS Position Itself? Where to Engage? To Be Determined

The elective affinities I have described in this essay present both sticky challenges and uncertain opportunities. As I have sought to show, at least in some spheres, data science is framed not only as a matter of technical capacity, resources, and skills but at times also concerned with variously defined organizational and social problems associated with data. Some of these problems are internal, or “functional” (Berg 1998), to the goals of data science—such as understanding the specificity of domains or problems of

interdisciplinary communication—but others are more classically societal, such as the downstream consequences of novel computational capacities and data analytics for privacy, access, or equity. This double problematization of the social offers multiple avenues for STS scholars to engage in the rollout of data science.

One facet of STS scholarship has engaged elements of this intersection with a focus on the consequences of, for instance, human interactional data, platforms and algorithms of social media, and other public-facing data-driven architectures (Gillespie 2010; boyd and Crawford 2012). Today, there is a great deal of razzle and dazzle at public intersections with data, with visible and proximate consequences following security and privacy breaches, emotional manipulations, or transformed and unregulated work. The concern with social media has been a superb “hook” for drawing attention to the importance of technical transformations in everyday lives. These critical understandings of the social are a hard-won and incomplete victory by those concerned with the downstream consequences of data. But often this line scholarship, as a genre, demands demonstration of close links between technicalities and *the public*: for example, social media users, or financial subjects.

The demand for a tight and immediate link between technical capacities and social (justice) consequences can be constraining for the investigations I am suggesting here, as those explicit links to social consequence may be nascent, speculative, or simply not present. As I have sought to show, social media and other social data represent only a slice of what data science targets across the natural and social sciences, as well as in the depths of (rather than public facing) activities of business and state. Perhaps more importantly, focusing only on social data or other data usages in direct proximity to publics fails to recognize the *transverse* ambitions of data science I have begun to outline here. It is in these transverse movements that we will find “where the action is” for data science, including the development of new computational capacities and tools, and experts with the growing skill sets to work across domains or to render the data and other materials of domains more readily available to make those crossings.

A central tenet of STS has been that the epistemological and ontological crafting occurring in science and technology is impactful even while distal to the public sphere, arguably *because* they are initially distant from the public sphere, for example, the lab (Latour and Woolgar 1979) or the experiment (Shapin and Schaffer 1985). The interstices I am pointing to here require more empirical and analytic work, and just the kind that STS scholars are eminently skilled to investigate with our penchant to gain the

interactional expertise necessary to make sense of technical, esoteric, and black-boxed human action and artifacts. Similarly, a central tenet of infra-structure studies has been that debates and controversies settled “deep” within sociotechnical organizations may have far reaching consequences as they circulate on the increasingly well-paved highways of interoperability (Ribes 2017): for example, algorithms, analytic styles, tools, data, and visualizations. An innovation in computerized molecular modeling from the 1980s may come to undergird contemporary crowdsourced analysis of proteomics in HIV research (Ribes et al. 2012). Tracing such pathways is a daunting task for any scholar. If data science is generating new kinds of transverse domain relationships, then to make sense of these, STS scholars will have to display an equal investigative agility by placing themselves at key intersections (whether new or old) and following the lateral movement of data, analytics, tools, and techniques across them.

Such lateral movements are often *already* framed as social challenges, for example, of collaboration, disciplinary difference, or organizational communication. Whether framed as “functional” (Berg 1998) contributors or through hard-won and sustained concerns about power, inclusion, or representation, social and humanistic scholars have long sought to carve and sustain niches for their contributions informing inquiry, design, maintenance, or policy. While certainly an incomplete project, such niches do offer ready-made avenues for STS engagement at the deep interstices of data science’s technological planning, development, and rollout. However, as configured today, those ready-made venues for “the social” come at a cost.

I say that these ready-made venues come at a cost because they position social scientists very particularly, often as a kind of “service science,” there to know the domains or smooth the complexity of collaboration. Another looming danger is reinscribing social consequences as distinct from technical and scientific concerns. In my experience, the ready-made niches for social science in data science come with ready-made expectations and boundaries for contribution or engagement. However, such roles are not fixed; if they are the “hard-won” outcomes of past efforts, there are new hard-won outcomes to be had. In this essay, I have shown continuities across CI and data science but also pointed to what seem to be new, emergent configurations, well beyond functional roles for “social science.” And there lie our opportunities.

Painting the broader emerging picture of data, technology, and expert flows will demand STS scholars positioned “in domains,” at the development of intersections “across domains,” as well as inspecting the activities at an emerging theoretical and technical core that professes to be domain independent even while aspiring to come in contact with all of the domains.

In addition to describing, representing, or theorizing data science, STS researchers have an opportunity to shape its rollout, whether informing design, doing it, or something else. And since the activities of data science will undoubtedly continue to overflow any technical definition, STS too will be entangled with the rollout of data science and its consequences.

Acknowledgments

The author would like to thank Andrew Hoffman, Sarah Inman, Charles Hahn, Steven Slota, Daniela Rosner, Geoffrey Bowker and Katie Vann for their comments on drafts.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported by the National Science Foundation Award #1638903, Institutionalizing the Data Sciences, a Sociotechnical Investigation of BDHubs.

Notes

1. The term precedes its social theoretical meaning, drawing inspiration from chemistry, and then a metaphorical transposition into literature (Howe 1978). Sociologist Andrew McKinnon (2010) has characterized elective affinities as “best understood as an analysis of emergence in the chemistry of social relations” (p. 108, emphasis added).
2. Cyberinfrastructure (CI) continues today but with somewhat more modest ambitions. It waned with the dissolution of the Office of Cyberinfrastructure at the National Science Foundation (NSF) in 2013 but continued as that office was first absorbed into the computer and information science directorate and thereafter has traced a winding trajectory across the NSF.
3. Occasionally, some subsection of computer science may be termed a domain—say, networking—thereafter to be served in some way by having access to the more general approaches or tools of data science.
4. It is notable that in collaborations of domain and computer scientists, technologists often face the same kind of challenges. They too chafe at receiving non-research requests such as being asked to develop websites or to clean data sets. Service is important and valuable; the expectation of service only becomes problematic relative to a person who identifies as doing something else, in this case novel research.

References

- Abbott, Andrew Delano. 1988. *The System of Professions: An Essay on the Division of Expert Labor*. Chicago, IL: University of Chicago Press.
- Atkins, Daniel E. (Chair). 2003. *Revolutionizing Science and Engineering through Cyberinfrastructure: Report of the National Science Foundation Blue-ribbon Advisory Panel on Cyberinfrastructure*. Washington, DC: National Science Foundation.
- Berg, Marc. 1998. "The Politics of Technology: On Bringing Social Theory into Technological Design." *Science, Technology, & Human Values* 23 (4): 456-90.
- Bergman, Mark, Kalle Lyytinen, and Gloria Mark. 2007. "Boundary Objects in Design: An Ecological View of Design Artifacts." *Journal of the Association for Information Systems* 8 (11): 546.
- Bowker, Geoffrey C. 1993. "How to Be Universal: Some Cybernetic Strategies, 1943-70." *Social Studies of Science* 23 (1): 107-27.
- Bowker, Geoffrey C., Stefan Timmermans, and S. L. Star. 1997. "Infrastructure and Organizational Transformation: Classifying Nurses' Work." In *Information Technology and Changes in Organizational Work*, edited by W. Orlikowski, G. Walsham, M. Jones, and J. DeGross, 344-70. London, UK: Chapman and Hall.
- boyd, Danah, and Kate Crawford. 2012. "Critical Questions for Big Data: Provocations for a Cultural, Technological, and Scholarly Phenomenon." *Information, Communication & Society* 15 (5): 662-79.
- CRA (Computing Research Association). 2016. "Computing Research and the Emerging Field of Data Science." Accessed July 7, 2017. <https://cra.org/data-science/>.
- Daston, Lorraine, ed. 2000. *Biographies of Scientific Objects*. Chicago, IL: University of Chicago Press.
- Dourish, Paul. 2001. *Where the Action Is: The Foundations of Embodied Interaction*. Cambridge, MA: MIT Press.
- Dourish, Paul. 2006. "Implications for Design." *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*:541-50.
- Dourish, Paul, and G. Button. 1998. "On 'Technomethodology': Foundational Relationships between Ethnomethodology and System Design." *Human Computer Interaction* 13 (5): 395-432.
- Edwards, P. N., S. J. Jackson, M. K. Chalmers, G. C Bowker, C. L. Borgman, D. Ribes, M. Burton, and S. Calvert. 2013. *Knowledge Infrastructures: Intellectual Frameworks and Research Challenges*. Ann Arbor, MI: Deep Blue.
- Epstein, Steven. 2007. *Inclusion: The Politics of Difference in Medical Research*. Chicago, IL: University of Chicago Press.
- Foucault, Michel. (1975) 1995. *Discipline and Punish*. New York: Vintage Books.

- Galison, Peter Louis. 1999. "Trading Zone: Coordinating Action and Belief." In *The Science Studies Reader*, edited by M. Biagioli, 137-60. New York: Routledge.
- Gieryn, Thomas F. 1983. "Boundary-work and the Demarcation of Science from Non-science: Strains and Interests in Professional Ideologies of Scientists." *American Sociological Review* 48 (6): 781-95.
- Gillespie, Tarleton. 2010. "The Politics of 'Platforms'." *New Media & Society* 12 (3): 347-64.
- Heims, Steve J. 1991. *The Cybernetics Group*. Cambridge, MA: MIT Press.
- Henderson, Kathryn. 1991. "Flexible Sketches and Inflexible Data Bases: Visual Communication, Conscription Devices, and Boundary Objects in Design Engineering." *Science, Technology, & Human Values* 16 (4): 448-73.
- Hjørland, Birger, and Hanne Albrechtsen. 1995. "Toward a New Horizon in Information Science: Domain-analysis." *Journal of the Association for Information Science and Technology* 46 (6): 400-425.
- Howe, Richard Herbert. 1978. "Max Weber's Elective Affinities: Sociology within the Bounds of Pure Reason." *American Journal of Sociology* 84 (2): 366-85. doi: 10.1086/226788.
- Jackson, S. J., P. N. Edwards, Geoffrey C. Bowker, and Cory P. Knobel. 2007. "Understanding Infrastructure: History, Heuristics, and Cyberinfrastructure Policy." *First Monday* 12 (6).
- Kjeldskov, Jesper, and Jeni Paay. 2005. "Just-for-us: A Context-aware Mobile Information System Facilitating Sociality." *Proceedings of the 7th International Conference on Human Computer Interaction with Mobile Devices & Services*: 23-30.
- Klein, J. T. 1996. *Crossing Boundaries: Knowledge, Disciplinarity and Interdisciplinarity*. Charlottesville: University Press Virginia.
- Knorr-Cetina, K. 1997. "Sociality with Objects: Social Relations in Postsocial Knowledge Societies." *Theory, Culture & Society* 14 (4): 1-30.
- Latour, Bruno. (1991) 1993. *We Have Never Been Modern*. Cambridge, MA: Harvard University Press.
- Latour, Bruno, and Steve Woolgar. 1979. *Laboratory Life: The Construction of Scientific Facts*, *Sage Library of Social Research*, Vol. 80. Beverly Hills, CA: Sage.
- Lee, Charlotte P., Paul Dourish, and Gloria Mark. 2006. "The Human Infrastructure of Cyberinfrastructure." *Proceedings of the 2006 Conference on Computer Supported Cooperative Work*: 483-92.
- Leonelli, Sabina, Brian Rappert, and Gail Davies. 2016. "Data Shadows: Knowledge, Openness, and Absence." *Science, Technology, & Human Values* 4 (2): 306-10.
- McKinnon, Andrew M. 2010. "Elective Affinities of the Protestant Ethic: Weber and the Chemistry of Capitalism*." *Sociological Theory* 28 (1): 108-26.

- Neighbors, James. 1980. "Software Construction Using Components." PhD diss., Department of Informatics, University of California, Irvine.
- Orr, Julian E. 1996. *Talking about Machines: An Ethnography of a Modern Job*. Ithaca, NY: Cornell University Press.
- Pollock, N., and R. Williams. 2008. *Software and Organizations: The Biography of the Enterprise-wide System or How SAP Conquered the World*. New York: Routledge.
- Ribes, David. 2014. "Ethnography of Scaling, or, How to Fit a National Research Infrastructure in the Room." *Computer Supported Cooperative Work (CSCW'14)*:158-70.
- Ribes, David. 2017. "Notes on the Concept of Data Interoperability: Cases from an Ecology of AIDS Research Infrastructures." *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*: 1514-26.
- Ribes, David, and K.S. Baker. 2007. "Modes of Social Science Engagement in Community Infrastructure Design." In *Proceedings of Third International Conference on Communities and Technology*, edited by Steinfield Brian T. Pentland, M. Ackerman, and N. Contractor, 107-30. London, UK: Springer.
- Ribes, David, and Geoffrey C. Bowker. 2009. "Between Meaning and Machine: Learning to Represent the Knowledge of Communities." *Information and Organization* 19 (4): 199-217.
- Ribes, David, and Thomas A. Finholt. 2008. "Representing Community: Knowing Users in the Face of Changing Constituencies." *Proceedings of the 2008 ACM Conference on Computer-Supported Cooperative Work (CSCW)*:107-16.
- Ribes, David, Steven J. Jackson, R. Stuart Geiger, Matthew Burton, and Thomas Finholt. 2012. "Artifacts that Organize: Delegation in the Distributed Organization." *Information and Organization* 23 (1): 1-14.
- Rose, Nikolas. 1990. *Governing the Soul: The Shaping of the Private Self*. London, UK: Routledge.
- Shapin, Steve, and Simon Schaffer. 1985. *Leviathan and the Air-pump: Hobbes, Boyle and the Experimental Life*. Princeton, NJ: Princeton University Press.
- Simon, Herbert A. 1980. "Cognitive Science: The Newest Science of the Artificial." *Cognitive Science* 4 (1): 33-46.
- Simone, Carla, Gloria Mark, and Dario Giubbilei. 1999. "Interoperability as Means of Articulation Work." *Proceeding of ACM Conference on Work Activities Coordination and Collaboration (WACC'99)*, February 22-25, San Francisco, CA.
- Slayton, Rebecca. 2013. *Arguments that Count: Physics, Computing, and Missile Defense, 1949-2012*. Cambridge, MA: MIT Press.

- Star, Susan Leigh. 1989. "The Structure of Ill-structured Solutions: Boundary Objects and Heterogeneous Distributed Problem Solving." *Distributed Artificial Intelligence* 1:37-54.
- Star, S. L. 2010. "This Is Not a Boundary Object: Reflections on the Origin of a Concept." *Science, Technology, & Human Values* 35 (5): 601-17.
- Star, S. L., and K. Ruhleder. 1994. "Steps towards and Ecology of Infrastructure: Complex Problems in Design and Access for Large-scale Collaborative Systems." *Proceedings of the Conference on Computer Supported Cooperative Work*: 253-64.
- Strathern, Marilyn. 1995. *The Relation: Issues in Complexity and Scale*. Cambridge, UK: Prickly Pear Press.
- Suchman, Lucy. 1987. *Plans and Situated Actions*. Cambridge, MA: Cambridge University Press.
- Szalay, Alexander, and Stuart Feldman. 2018. "Spreading the Use of Big Data in Scientific Research." Accessed August 10, 2018. https://www.nsf.gov/events/event_summ.jsp?cntn_id=296243&WT.mc_id=USNSF_13&WT.mc_ev=click.
- Vertesi, Janet, David Ribes, Laura Forlano, Yanni Loukissas, and Marisa Leavitt Cohn. 2016. "Engaging, Designing, and Making Digital Systems." In *The Handbook of Science and Technology Studies*, Fourth Edition, edited by Felt Ulrike, Fouché Rayvon, A. Clark Miller, and Laurel Smith-Doerr, 169-94. Cambridge, MA: MIT Press.
- Weber, Max. 2013 [1904]. *The Protestant ethic and the spirit of capitalism*. New York, NY: Routledge.
- Williams, Raymond. 2014 [1976]. *Keywords: A Vocabulary of Culture and Society*. Oxford, UK: Oxford University Press.

Author Biography

David Ribes is associate professor in the Department of Human Centered Design and Engineering (HCDE) and director of the Data Ecologies Lab (deLAB) at the University of Washington. He is a sociologist of science and technology 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, transformations in objects of research. Please see davidribes.com or dataecologi.es for more.