Engaging Science, Technology, and Society 6 (2020), 111-132 DOI:10.17351/ests2020.341

Labor Out of Place:
On the Varieties and Valences of (In)visible Labor in Data-Intensive Science

MICHAEL J. SCROGGINS[1]
UNIVERSITY OF CALIFORNIA, LOS ANGELES

IRENE V. PASQUETTO[2]
HARVARD KENNEDY SCHOOL

Abstract
We apply the concept of invisible labor, as developed by labor scholars over the last forty years, to data-intensive science. Drawing on a fifteen-year corpus of research into multiple domains of data-intensive science, we use a series of ethnographic vignettes to offer a snapshot of the varieties and valences of labor in data-intensive science. We conceptualize data-intensive science as an evolving field and set of practices and highlight parallels between the labor literature and Science and Technology Studies. Further, we note where data-intensive science intersects and overlaps with broader trends in the 21st century economy. In closing, we argue for further research that takes scientific work and labor as its starting point.

Keywords
data-intensive science; invisible labor

Introduction
For Vannevar Bush, 1945 was an annus mirabilis. His wartime efforts at the Office of Scientific Research and Development (OSRD) to reorganize American science along lines established by industrial engineering had borne fruit in the form of scientific advancements and overflowing archives. In chemistry and physics, Bush found a model of scientific practice radically changed from the prewar era; research conducted by teams of cooperating scientists, intensive publishing schedules, and wide dissemination of research findings became the new normal. This new model of science, Bush argued, could and should be expanded to the whole of science. Cementing these changes would be a wholesale change in the organization and funding of American science (see Mirowski 2011), and much of the OSRD model would find its way into the postwar National

[1]Michael J. Scroggins, Email: michaeljscroggins@ucla.edu
[2]Irene V. Pasquetto, Email: irene_pasquetto@hks.harvard.edu

Science Foundation (NSF) along lines laid out by Bush in *Science: The Endless Frontier* (Bush 1945a). One direction of Bush's vision called for inculcating a new set of sentiments within scientific apprentices. In contrast to their prewar counterparts, scientists trained in the postwar era needed to be comfortable with cooperatively communicating and disseminating research, working at the faster wartime pace normalized in physics and chemistry, and bridging the gaps occasioned by increasing disciplinary specialization. Another direction called for the flood of information produced by this new model of science to be tamed through technology and automation, such as Bush's own Memex (Bush 1945b).

Two decades after Bush's annus mirabilis, his speculation that the onerous work of science could be automated through technological advance ran headlong into a series of inconvenient facts. In a 1961 provocation on the monumental scale of Big Science and its financial demands, Alvin M. Weinberg observed in the journal *Science*, "if [a scientist] becomes involved with Big Science [the scientist] will have to become a publicist, if not a journalist, an administrator, as well as a spender of big money" (Weinberg 1961, 161). Weinberg lamented the new distribution of scientific practiced assumed by Bush's vision, deliberated in the popular press and congressional hearings rather than in lecture halls, and the celebration of spectacular and dramatic research at the expense of the perceptive and mundane. A few years later, in a speech to the 1968 American Psychological Association titled "As We May Think, Information Systems Do Not," William Paisley observed that despite the computational power, technology, and data available in 1968, the archives had failed to automate: "the missing element in information service is people. Mediators. Middlemen… our dreams for mechanizing the archives and making them truly responsive to researches and other users are dreams for the next century" (Paisley 1968, 13).

Science, as Weinberg and Paisley remind us, is a conversation between and betwixt collaborators and interlocutors (be they human, animal, mineral, mechanical, or digital), not a monologue given by a scientist to an army of functionaries. Fifty years on, their observations are no less true. We write in the midst of a transition from an era of industrial organization and Big Science to an era of data-driven science and Big Data. Like all technological transitions, there is both change and continuity in this transition. While theorizing the changes and continuities between Big Science and data-intensive science is beyond the scope of this article, we make common cause with Weinberg and Paisley by offering a critique and correction to the optimistic assumptions that have accompanied data-intensive science. Our parenthetical expression (in)visible labor draws attention to the stakes of our inquiry. While invisible labor has long been a part of post-war science, and a concern for critical voices, data-intensive science has been accompanied by distinctive forms of labor and inquiry. Behind data-intensive science's technological facade lies a bewildering array of human labor, some performed in the spotlight by star scientists, but most performed behind the scenes by the precariously employed in conjunction with computational machines.

In what follows we draw on an archive comprising fifteen years of continuous and comparative research across multiple domains of data-intensive science to present a bottom-up snapshot of the (in)visible labor engendered by data-intensive science. Initially, we use the emerging literature on data-intensive science and classic works in science studies to delineate a stock of activities necessary and common to scientific practice: authoring, administering, maintaining, archiving, and collaborating. We use labor literature and ethnographic vignettes drawn from our data corpus to illustrate how data-intensive science has animated new forms of scientific labor, many foreshadowed by emerging forms of work in industry. Our analysis also highlights moments of continuity and disruption between invisible labor before and after the advent of data-intensive science. Borrowing a chemical and grammatical term, valence, we show how data-intensive science combines and binds the common activities of Big Science to new forms of work and labor, in both familiar and surprising ways.

What Makes a Science Data-Intensive?

By data-intensive science we mean scientific fields in which the quantity and velocity of data generation have led to (i) a reliance on computational power and techniques to analyze, curate, and archive data (Kitchin 2014a, 2014b, 5–7; Burns, Vogelstein, and Szalay 2014; Critchlow and Dam 2013; Ekbia et al. 2015); and (ii) intensifying transdisciplinary collaboration (Ribes and Bowker 2008; Bell, Hey, and Szalay 2009). Data-intensive science is heterogeneous in composition and complex in operation. The distinguishing characteristic of data-intensive science is the intensive production, consumption, and circulation of data and data products. Digital data is the sine non qua of data-intensive science, and analog data must be converted (often laboriously) to digital form for manipulation, analysis, circulation, and storage. As a result, the cardinal methods of analysis in data-intensive science are practiced on a screen using common digital tools such as spreadsheets, databases, and code editors. Adding to this complicated environment is the capital-intensive nature of data-intensive science, typically administered through large grant-funded projects with budgets in the tens or hundreds of millions that require an extra layer of administration, public outreach, and data management. Because of the large budgets and long life of the resultant infrastructure, data-intensive science requires intensive interdisciplinary collaborations that must be maintained through time and across space.

As Bush played a key role in the transformation of science in the postwar period from his perch at the OSRD, another engineer, Jim Gray, played a key role in ushering in the era of data-intensive science from his position at Microsoft Corporation. As Bush looked to physics and chemistry as models for scientific practice, Gray looked to astronomy as a model for data-intensive science. In a lecture summarizing his long-term collaboration with astronomers (Hey, Tansley, and Tolle 2009), Gray articulated his vision for a 4th scientific paradigm based on the intensive generation, analysis, and archiving of scientific information—from lab books to grey literature to data sets—stored in databases and connected through the internet. Once enough data is accumulated, per Gray, the main analytic focus will naturally shift from the analysis of bespoke data sets to the reanalysis of extant data sets generated automatically and made publicly available. While Gray's vision has not been universally adopted (the humanities and field sciences are notable holdouts while large parts of astronomy have adopted Gray's vision), in disciplines where data-intensive science has become the normative mode, new forms of labor, notably analyzing and archiving large data sets, have emerged in its wake.

Yet, despite Gray's rhetoric of a 4th scientific paradigm on the horizon, data-intensive science has not caused a qualitative break with past scientific practice. Data-intensive science is, perhaps, best understood as an amalgamation of traditional scientific practice overlaid with elements of computer-supported work as practiced in industry and archiving practices derived from the development of relational databases (Gray 1996).

Research Design

The analysis presented here is based on interviews and ethnographic observations conducted by successive cohorts of researchers at the UCLA Center for Knowledge Infrastructures (henceforth CKI), including the authors of this paper, over a fifteen-year period beginning in 2004. The authors have contributed research and analysis to the CKI's ongoing research program since 2014. Studies were conducted in a diverse array of scientific fields, specializations, and sectors across the physical, life, and social sciences (Mayernik, Wallis, and Borgman 2013; Wallis, Rolando, and Borgman 2013; Darch et al. 2015; Borgman et al. 2015; Pasquetto, Randles, and Borgman 2017; Pasquetto, Borgman, and Wofford 2019; Souleles and Scroggins 2017; Scroggins 2017, 2019). Research participants include PIs, postdoctoral researchers, doctoral students, graduate students, technicians, librarians, and staff.

Ethnographic research incorporated in this article includes field observations of participants performing research, laboratory and community meetings, and other events. Interviews were audio recorded, transcribed, and complemented by the interviewers' memos on noteworthy topics and themes. In sum, the CKI's corpus consists of 483 coded interview transcripts, an equal number of ethnographic observations, and thousands of pages of documents, listserv archives, and other grey literature.

Though sharing a common orientation towards the intensive use of data, data-intensive science is not a monolith. Differences in tools and work practices, disciplinary concerns, epistemological and ontological assumptions, the scale and centrality of data generation, and the state of standards and measurements within a given discipline all play their part in the kinds and amounts of labor required. For example, astronomy, with standardized file formats and common instruments (telescopes) with consistent metrology, requires different forms and amounts of labor than ecology, which lacks standardized file formats, common instruments, and consistent metrology. Though both astronomy and ecology use data intensively, they produce, consume, and circulate data in differing manners. To account for this variation, in our reanalysis of the CKI corpus we created an extensive list of activities, some discipline-independent and others discipline-specific, related to work and labor.

From our list of activities, we then developed a set of indexes and codes in ATLAS.ti. Next, using our codes, we drew a selection of ethnographic vignettes highlighting the contours and boundaries of these ongoing activities. By making activities, rather than job classifications or place in the academic hierarchy, central to our analysis we demonstrate how the everyday work of science cuts across received classifications and hierarchies. These activities form the bones of our analysis, with ethnographic vignettes providing the flesh. Confidentiality permitting, where possible we have allowed our research participants to speak in their own voices.

Opening the Black Box of (In)visible Labor in Data-Intensive Science

Invisible labor was canonically defined by Daniels (1987) as work "that disappears from our observations and reckonings." The reckonings from which invisible work disappears are classificatory schemes and metrics, such as job descriptions and economic indicators, that make work and labor visible and rewardable to and within institutions (Star and Strauss 1999; Bowker and Star 1999). Daniels' remarks summarize a body of work inaugurated by feminist scholars in the 1970s, when feminist scholars began to look critically at the relationship between labor performed inside the household and labor performed outside the household. A number of researchers and theorists found labor outside the house, i.e. labor reflected in rewards structures and economic indicators, was reliant on an equal, and often greater, amount of invisible, unrewarded, and gendered labor performed within households (Fee 1976; Himmelweit and Mohun 1977; Coulson, Maga, and Wainwright 1975). An important point for thinking through scientific practice is Cowan's (1976) demonstration that the forms invisible labor assumes cannot be separated from the wider political economy in which it is embedded.

In the years since Daniels' 1987 comments, the visibility/invisibility dichotomy has acquired more nuances as scholars have (i) stressed that visibility is a multivalent and multifaceted spectrum, one with fractal folds and complexities and (ii) sought increasingly fine-grained and nuanced concepts to account for the complexities of labor in the 21st century. By the 1980s, as the economy shifted from its industrial base towards the service sector, scholars such as Hochschild (2012) conceptualized emotional labor as a new type of invisible labor: the transformation of domestic work into a commodity necessary to the functioning of the service economy. Following on Hochschild's pioneering work, the concept of care has emerged as a critical standpoint from which to examine labor in all its manifested forms (Star 1990; Mol 2008; Tronto 1993), but particularly the labor of maintaining human relationships. Tronto defined care as "all we do to continue, maintain, and repair 'our world' so that we can live in it as well as possible" (1993). While care has been widely taken up as an ethical stance in terms of human relationships, Jackson (2017) emphasized that care also has material implications. In data-intensive science, care is important as both an ethical stance towards both the repair and maintenance of scientific machinery and the human relationships that animate scientific inquiry.

Within history of science, invisible labor has often been conceptualized as unseen, under-rewarded, gendered, and classed labor accompanying the work of the scientist. Shapin's (1989) description of the invisible labor in Boyle's laboratory is the classic account, with recent accounts of the gendered and intersectional complexity of scientific work and theory (Roberts 2018; Harding 2016) carrying this work forward into contemporary scientific practice. While we do not directly address the work of funding data-intensive projects, the necessity of seeking grant-based funding within a marketplace of competing projects forms the political economy of 21st century science (Lave, Mirowski, and Randalls 2010; Tyfield 2013; Edgerton 2017) and creates the context within which labor acquires (in)visibility. On one side of the ledger is labor that contributes to winning grants, and on the other side is labor that "disappears from our observations and reckonings" when grant applications, tenure cases, publications, and public relations are given pride of place.

In the digital domain, the ethereal nature of digital work creates distinctive difficulties determining whose labor should be measured in official metrics, be they the mythical "man hours" in software development or scientific papers written and grants won, and whose labor is symbiotic to the metric. For example, Nardi and Engeström (1999) conceptualized digital work as a "web on the wind," structured in practice yet lacking the permanence required for institutional recognition. Building on the earlier scholarship in Computer Supported Cooperative Work and Human Computer Interaction is a thread of research into the difficulties of collaboration in digital environments. Though technological tools can sometimes be used as the main medium of collaboration, more commonly, personal relationships and an ethic of care must carry the collaboration forward, particularly in the case of transdisciplinary collaborations (Ribes and Bowker 2008). Similarly, Leonelli (2010) and Plantin (2018) have argued that in data-intensive science, the work of cleaning data sets, such as assigning metadata, developing ontological schemes, and checking instrument readings, is typically devalued as "non scientific work," yet both Leonelli's (2010) research on biocurators' work and Plantin's (2018) on "data cleaners" in a social science archive demonstrate that the creation of ontologies and metadata and the cleaning and documenting of data sets, respectively, is essential for data-intensive science.

While the canonical definition and foundational works on invisible labor emphasize the negative ramification of invisibility, two works (Orr 1996; Allen 2014) focusing on professional labor implicitly argue that not all invisible labor is negative. Work performed by nurses within a hospital is invisible in the canonical sense, but Allen (2014) suggested that the demands of professional practice in environments where sensitive information is archived, reused, and deliberated over can make a virtue of invisibility. Allen found that nurses, as part of their invisible yet essential professional practice, work across the boundary separating the formal elements of hospital care, such as scheduling medications, assigning beds, consulting with doctors, and filling out paperwork, from the informal aspects of hospital care, such as soothing patients and family and discussing cases and colleagues with fellow nurses. This work is essential to care but also to patient privacy. Likewise, in a study of Xerox copier repair technicians, Orr (1996) found the repair technicians' talk about machines, customers, and salespeople, more so than technical ability, to be their primary form of labor. Talk was used by the technicians to train apprentices, understand how and why copiers break down, and manage customer's and sales manager's expectations. In both instances, the professional talk of nurses and repair technicians police and repair the boundary between customer and company and carve out a class of professionals whose job becomes translating across that boundary. In data-intensive science, the professional work of science qua science occurs in negotiations over authorship credit, discussing the future of controversial lines of research, and mentoring apprentices. In sensitive contexts, the invisibility of professional talk can become a form of care.

If invisible labor is that labor which "disappears from our observations and reckonings," then visible labor is the unmarked category of labor made visible through classificatory schemes (Bowker and Star 1999; 2000). Writing for publication is the canonical visible work of the scientist. The work that technicians, administrators, and staff perform according to the standards negotiated in job classifications, descriptions, and annual reviews can be considered invisible. New standards bring new forms of visible work. For example, publishing data sets and research software has quietly become part of the visible work of science in some fields, as granting agencies have begun requiring data used in writing papers to be made publicly available (National Institutes of Health 2017).

Within the labor literature, hypervisible labor (Crain, Poster, and Cherry 2016) has emerged as a category denoting labor in which the aesthetic enjoyment of observing work, as a skillful achievement or spectacular failure, is the main attraction. The work of celebrity chefs in open kitchens, actors on a stage or screen, and servers at theme restaurants are all examples of hypervisible work. Like all work, hypervisible work carries both positive and negative ramifications. University donors attending academic conferences or spending time "on the mountain" with astronomers or in the lab with scientists are paying to experience the hypervisible work qua work of science. Work featured in the popular press, the work of public intellectuals, work with strong and direct ties to areas of policy, and work of interest to major foundations and donors we also consider hypervisible. On the other hand, scientific scandals and controversial research may become hypervisible in a negative manner. The other side of credit for success is blame for failures. An example of negative hypervisible scientific work is the recent spate of "outings" over the difficulty of reproducing studies in social psychology and other fields (Dominus 2017; Marcus and Oransky 2018).

The Varieties and Valences of (In)visible Labor in Data-Intensive Science

In the following section we employ vignettes to explicate the valences of (in)visible labor in data-intensive science in service to understanding the terrain of this quickly evolving scientific paradigm. We highlight the invisible, visible, and professional labor of data-intensive science in its complex permutations. Though our corpus contains several instances of hypervisible labor, in the interest of confidentiality we have excluded them from the following vignettes.

Authoring

By authoring we intend the work Foucault (2012) glossed as enunciating the truths of science. Unlike the singular authors of literary and artistic works, scientific authorship is plural. No one person, or machine, has full claim to the truths of nature but many have partial claim. Today, those partial truths are spread among instruments, papers, data sets, and software. *In data-intensive science, authoring* is both writing papers and grants and authoring data sets and research software (Mayernik et al. 2015; Green 2009; Hills et al. 2015). Authorship is also a dividing line separating those whose name appears on the byline of scientific papers, and hence registered as a citation, from those merely acknowledged. Adding to the complexity, in disciplines such as astronomy, it is customary to place the instrument that took the observations on the byline. A contested form of authoring specific to data-intensive science is creating datasets and research software. Data sets are often authored by graduate students, postdoctoral researchers, or technicians. Despite being central to data-intensive science and playing a key role in ensuring that data are able to circulate and remain interpretable, data and software authorships goes uncredited in publications and uncounted in tenure cases (Howison and Bullard 2016; Velden et al. 2014). Below we present three vignettes drawn from our data corpus that illustrate the valences of authorship in data-intensive science.

When postdoctoral researchers from different disciplines are joint authors, professional labor is required to assign credit: Authorship norms vary between disciplines, and negotiating the authorship order between authors hailing from different disciplines, in this case physics and astronomy, requires balancing the sometimes competing claims of the work on the paper against disciplinary norms for distributing credit. This is professional labor proceeding, as Allen (2014) and Orr (1996) observed, through talk and negotiation. One participant described the delicate dance of arranging an article's byline according to the field in which a postdoctoral researcher is applying for a job:

> It was a negotiation primarily between the particle physicist and the astronomers…we can't just ignore the fact that astronomy departments have different criteria than physics departments about when people are applying for faculty jobs. We have to accommodate that and since the particle physicists don't care who's first author, we say, "All right, let the astronomers be first author." ... The way it works in particle physics you might say, "How do particle physicists ever get faculty jobs?" The answer is you rely on letters from the people in the collaboration who know what they did. So the procedures are just significantly different. It requires, it's a delicate thing, how do you balance the needs of the two different groups.

Technicians' labor is not always invisible; at the PI's discretion technicians can be authors: Data-intensive science has not changed the scientific hierarchy; a technician's inclusion on the byline is determined by the judgment of the PI, as a project's PI owns the project's data. The PI's judgment is especially important when a potential co-author has moved on from the research team, lacks the academic credentials to justify authorship, or their contribution is unclear. Here a postdoctoral researcher explains the decision to include a laboratory technician on the byline of a recent paper: "He was actually their lab technician, so he did a majority of that particular work. And, of course, he'll be included on the paper because he did... [the PI] is very good about, at least worst case scenario, making sure you get credit for what you did because they're not here to defend themselves."

Formerly invisible authors of datasets can be made visible, with caveats: Cleaned and archived data sets are the lifeblood of data-intensive science. Yet, cleaning and archiving data sets is often uncredited and unrewarded work within scientific disciplines. One optimistic possibility is an emerging technique of data publishing and authorship used in some astronomical projects that grants authorship credit to all those, no matter their role or status, who had a hand in producing a dataset. As one of our interviewees explains,

> Then you'll find at least one and sometimes two alphabetical tiers of authors which are indicating people whose work was not devoted to that particular science effort, but they've earned their authorship rights by virtue of helping enable the survey by doing kind of broader infrastructure work and made that work possible. And so they get what's called "architect status."

Data-intensive science has not changed the core difficulties inherent in scientific authorship. Deciding who among a plurality of authors (a problem that extends even to scientific instruments) deserves which share of the credit, and, therefore, visibility, is still an issue for debate and deliberation. Yet, one area where data-intensive science is slowly refashioning authorship norms is through the infrequent inclusion of software and dataset authors on the byline.

Administering

Administering, from the Latin *administrare*, means to "to help, assist, manage, control, guide." Administering is not ordinarily visible to those outside the project, and many aspects of administration are invisible to non-administrators inside a given project. Though not specific to data-intensive science, the labor of administering is essential to its course and conduct, paving the way for future rounds of funding by ensuring that grant requirements are met, reports are written, and outreach efforts are undertaken. Often this means taking responsibility for so-called softer, nonscientific parts of a large grant—education and diversity requirements, for example—and satisfying the demands of IRB boards, data privacy and security policies, and HR mandates. For staff, administering often requires developing interactional expertise, being able to talk authoritatively about a scientific domain (Collins and Evans 2007), while for faculty, PIs, and research assistants administering often means developing contributory expertise (Collins and Evans 2007) in administration, learning to contribute to the state-of-the-art project management. Below we present three vignettes illustrating the complexities of administering in data-intensive science.

Administrative assistants labor invisibly, doing the housekeeping of data-intensive science: The labor of administering is often spread across multiple positions. Official organizational charts that simply divide academic from nonacademic staff are of little help in discerning who administers what and how. One administrative assistant described the experience:

> I am an administrative assistant. I work on, mainly, event coordination and event planning. I'm also responsible for development of the center's website, and doing the cyberinfrastructure research to advise the Executive Committee, and then, sort of, everything else under the sun from administrative paperwork to event setup. We're sort of a catch-all kind of job… when I applied to this job, there was nothing in the job

> description about what sort of organizations this was or what they did… [I am] figuring out how to combine the administrative and the scientific databases with the public website, so establishing a website with data portals of various kinds to sort and to filter different entities based on their metadata.

Administrators do the visible labor of meeting metrics required by funders: The broader impacts of scientific work, important for securing funding but difficult to institute in practice, are often the responsibility of administrative staff on large projects. Particularly in projects with educational and diversity goals, administrators are often at the leading edge of creating new pathways into scientific careers for underserved communities. One project coordinator explained how she organized a gender equity program:

> One of our big things that we're also looking at was gender equity, especially in computer sciences... Sort of trying to demystify some of the stereotypes that revolve around women and the sciences.... Our hope was that we could educate people and provide enough positive experience for the outcomes to be that people feel as if there are pathways for making that more equal. But at the same token also, we were also gauging perception of what people felt. What is [sic] the capabilities, I suppose, or stereotypes of women in the sciences.

Administrators must be autodidacts, educating themselves in the professional labor of science: Also common is the autodidact administrator who, once in the position, must teach themselves enough domain science to be help, manage, assist, or guide a project. Another administrative assistant describes the struggle of coming up to speed with a fast moving scientific field:

> [The first day] was like, "Okay, you're the Education and Diversity Director. Here's what the grant said. Here's what this document says. Go for it."…[the PI] told me once that she considered professional development to be the ultimate sign of an employee being able to figure out what it was that they needed to do and go and do it. And when I came in, I had no background in science. So, I recognized that as the point of my greatest learning curve. And I began to attend the weekly lab meetings. Nobody made me go. Nobody suggested that I go. I just said, "I've got to go hear these students talk." So, they would share and I would ask questions. And frequently, I would talk to the grad students or the postdocs and afterwards, I'd make lists. I come back to my office. I Google the words that I had down on paper.

While data-intensive science has not significantly changed the form and content of administering labor, it has changed its tempo and scope. Data-intensive projects move at a faster speed and the collaborations they engender are wider and more transdisciplinary than scientific collaborations of the past. In response, administrative work has expanded into new areas and their work has acquired, at least in some contexts, increased visibility.

Maintaining

While the innovative analytical tools of data-intensive science have garnered the limelight (and funding), it is routine maintenance that makes science possible (Russell and Vinsel 2018). Like all contemporary science, data-intensive science requires a dazzling array of technical skills, most of which require ongoing education, both formal and informal, to master. Expertise with computational analysis and data cleaning is a particular requirement of data-intensive science, but technicians who build instruments and maintain equipment (often analogue) are also required. Technicians range from graduate students, who know a little more Python than anyone else, to electrical and software engineers with decades of specialized domain knowledge. Their jobs range from building and maintaining expensive instruments designed for decades of use, such as telescopes, to calibrating environmental sensors, to maintaining software codebases, to repairing all of the above and then some. Not until key components break, fall out of calibration, or fail to be constructed on schedule, does maintenance work come to the forefront. Yet, like all scientific technicians, the technicians of data-intensive science, no matter the importance of their contribution, are often in precarious positions and paid with the least stable forms of funding.

The invisible labor of technicians eases the PI's managerial burdens: The need to maintain equipment is a constant companion of all science, not just data-intensive science. Many scientists we have interviewed, such as the one quoted below, have drawn a surprisingly old distinction (Shapin 1989; Morus 2016) between technical staff and scientists over their relationship to equipment. In this case, a PI positively describes the role technicians played at a former institution and laments the additional labor required of him at his current institution:

> [There] you have a person for everything. You have a person who does orders. You have a person that if you cannot order something via the net or whatever, that person actually drives around to buy stuff. You have computer people that help with everything computer. You have technicians like mechanical and electronic technicians that help with all kinds of equipment. Here, we don't have that.

Students and postdoctoral researchers do the invisible work of maintaining analytic pipelines: A technician is also a *bricoleur*, skilled at combining the odds and ends of various systems and infrastructures into a workable whole. Here a PI describes how a data stream originating from a robot is rendered useful for analysis:

> So the problems are that it requires a fair amount personal intervention in the sense that I could draw a nice picture of this data flow but it's not anywhere near as automated as anyone might believe it is, if it involves typically... In this case it involved [a graduate student] doing a fair number of things manually… It's not automated, it's just enough to do experiments of this kind, but it doesn't really translate into a system or anything like that that we could just give to other people to do, right? So, and all the tools to massage the data once it comes off the robot are all sort of home brewed, right? And they tend to live because two or three students or postdocs sort of maintain them. But they aren't systematically maintained or archived or sort of curated in any way.

The metrology of large infrastructure is maintained through the invisible labor of skilled technicians: In astronomy and physics, the cost of physical infrastructure is measured in the hundreds of millions of dollars and is intended to serve thousands of scientists over several decades. To hold the measurement standards of such complex scientific instruments steady and thereby ensure the accuracy of the derived data, specially trained technicians must assist in the operation. A technician at an observatory explains the process:

> There are only two other telescopes built like this and we did this for a number of different reasons, primarily, to maintain what's called laminar flow between the optical elements primary and secondary mirror, to maintain image quality and temperature control…We have a crew that operates it, we collect the data, we give the data away, and you'll find that out that basically, we're a factory. We produce the data; we give it to people to use it… I do the mechanical, [another technician] does sort of the software side of things…then there's a series of technicians that work [in operations]. One electronic tech, one mechanical... and then a series of pluggers [technicians who press optical fibers into metal plates drilled with holes representing areas of interest in the night sky. This process makes astronomical observations usable in spectroscopic analysis.]. So we have like three pluggers, and what that means is these are the people who actually plug the plates during the day for observation purposes at night and that's pretty much the crew.

Maintenance is fundamental to scientific inquiry. Data-intensive science has added a digital overlay to existing maintenance requirements; adding datasets, analytical pipelines, archives, and software to the list of scientific instruments to be maintained and calibrated. As well, the digitization of so much data has seen scientists and their apprentices take on ad hoc maintenance and repair tasks, primarily when increasingly complex analytical pipelines breakdown in the midst of analysis.

Archiving

Archiving is the work of cleaning, wrangling, curating, and preserving data for future reuse. Archiving the digital data and data products generated by scientific instruments, or made digital by technicians, is the specific form of labor that makes data-intensive science identifiable as data-intensive. The technical and fiscal cost of generating data has fallen over the last few generations but the asymmetry Paisley identified when he observed that "information specialists" were needed to mediate between data banks and research groups has only grown (Moore et al. 2010; Mauthner, Parry, and Backett-Milburn 1998). Further adding to the complexity is the velocity at which data is generated, making the process of fixing data in place for archiving difficult. Despite these complexities, archiving is the lynchpin of data-intensive science. Without comprehensive archives full of papers and data sets with comprehensive and correct metadata, often a laborious and tedious process, the ease of analysis promised by digital workflows and pipelines disappears (Edwards et al. 2011).

The professional labor of science is often at odds with the professional labor of archivists: Archiving in data-intensive science means addressing the collision between file formats suited for cutting-edge research and file formats suited for archiving and preservation. Though astronomy is the rare discipline where one file format, FITS, predominates, localization of the file headers can cause problems for archivists, who need fixity to preserve data over the long term. An astronomical archivist working with a ground-based telescope explains:

> We didn't necessarily want to force [astronomers] into a standardization, because it tends to quell innovation and cleverness, and things like that… But then, we get different operational software or different detectors that collect the data in different ways, and, so all these [file] headers are slightly different from instrument to instrument and from era to era. That's one of our problems.

Behind the dream of automated data generation is the invisible labor of cleaning and munging data sets: Another common difficulty is overcoming the fragility of automatically generated data. Automation can save labor but automation can also be a cause of additional labor. The following vignette describes the case of a malfunctioning environmental sensor that caused malformed data to be automatically generated, necessitating manual cleaning:

> When we collected the data in Bangladesh, we had this really ad hoc way of saving the data…it's just what some guy came up with when he wrote the software. … [When the sensor malfunctioned] I had to spend hours and hours just cleaning the data because we had duplicate packets that were shown, data was printed out of order. And because of the software it was really hard to get it back in order. So if a node rebooted, any kind of node in the network rebooted, then the time stamps were screwed up, the sequence numbers were screwed up…I had to use anecdotal information, like okay, I know that I rebooted this node at this time…I did a lot of manual, like I think this time stamp should actually be this, so a lot of writing scripts to manually set time stamps, which never feels good.

Authoring metadata is invisible work that renders scientific papers and datasets visible: If authoring is the visible work of science, authoring the metadata that make data discoverable and usable is its invisible accompaniment. Compared to excitement of a published article or winning a grant proposal, success in archiving is decidedly unglamorous and made difficult by the emphasis on publishing over preservation. As an astronomical archivist commented,

> You want some kind of connection that's permanent, so when you see this link in whatever form it takes, it's gonna be good 20 years from now. It's that technical challenge…But there's also a social problem… People submit their papers and they don't provide the links [to the underlying data]. Partly it's culture, that's how astronomy has always been done. But there's another reason which is a little more self-serving, which is astronomers do not want to give other astronomers a leg up when doing research…There's also another difference, a big difference in the nature of data now and 20 years ago. Data used to be really expensive, and therefore people protected them.

> Data are now cheap, people don't need to protect them as much. This is the age of the big sky surveys, which are supported by very fast high quality computers that are capable of managing and processing huge amounts of data.

The archive is the starting and ending point for data-intensive science. Analysis starts with data found in an archive and data generation and collection end with data deposited in an archive. The centrality of archiving to data-intensive science has introduced new kinds of invisible labor in the form of authoring metadata and preparing datasets for long-term access. More cogently, archiving has brought conflicts between the norms of scientific practice and the norms of archiving practice as the standardization and consistency required of archival quality datasets is often at odds with the improvised and partial datasets common to the everyday work of science.

Collaborating

In data-intensive science, data, storage, and computing power take pride of place. But human relationships, mediated through data, storage, and computing power, produce scientific knowledge. And unlike computers, the human relationships that hold transdisciplinary research together require constant care and attention. Much like administering, the work of collaborating is not specific to data-intensive science. Collaboration has always been integral to scientific practice. Yet, the geographic dispersion of scientific collaborators, which is enabled and mediated by digital/computational methods and often formed by assembling or sharing common data sets, and the intense pressure towards interdisciplinary work have brought their own challenges to collaboration.

In this context we refer to emotional labor (Hochschild 2012) and relationship building as collaborating. Collaborating is an active verb, expressing action as well as attitude and ethical orientation. As Jackson (2017) has argued, the work of caring and cultivating an ethic of care has both symbolic and material dimensions and can be directed towards human relationships or relationships to materials and equipment that mediate between human relationships.

Administrators often do the emotional labor of counseling and mentoring students: One participant described an administrator who took on the emotional spadework required to make good on claims of furthering diversity and career formation:

> One of the things that I think was dramatically underappreciated by the management at large is the critical role that [the administrator] has played as a big brother, as a mentor towards lots and lots of students. And every program needs someone like him, whoever the official role is. As a person, he has been critical to the success of the center…. one of the remarkable sociological things about the center was that, there was probably at one point that half of the full-time staff, whether we were administrative or technical or whatever, were gay. For this department, engineering and so on, that was, I think, a quiet watershed event…in a very behind-the-scenes way and out-of-working-hours way, was somebody that many of our male students, at least gravitated to, to just be able to deal

> with that side of their lives. None of this ever was above the surface. There were never any gay bashing issues.... It was all very professional sort of a thing, but I think [he] has just had this incredible role.

PIs often do the emotional work of holding research teams together: A common refrain in long-term research teams is the PI who does the emotional spadework of making sure each member of the research team feels valued and appreciated. In an economic climate where many researchers and scientists work on short-term, grant-dependent contracts in precarious positions, the emotional work of making everyone feel valued is important to retaining key personnel. Here a research scientist on a short-term contract talks about of a PI who went out of her way to create the social conditions required for successful collaborations in data-intensive science: "One of things that I've appreciated about her is that more than any of the other faculty that I've worked at the School of Engineering, she has a degree of caring about people at a personal level that was very refreshing and rewarding."

Research scientists on temporary contracts often do the emotional work of bringing peers into conversation: One of our interviewees worked as a mediator, building trust and friendliness while avoiding "flame wars" between colleagues from the same discipline who nevertheless held divergent viewpoints on method, research, and analysis. Here the interviewee describes the difficulties of reshaping relationships built on competition into relationships built on cooperation:

> There were some really harsh emails, people didn't hold back on being critical of one another, and again, it just didn't help in terms of trying to build a cohesive team that was …we'd take people from two institutions, and we'd sit them next to each other in the same room, and it's like, "we're working together guys." So it was really challenging…and even if at the end of the day they still didn't fully respect each other or there was still some mistrust, I think I was able to develop a sense of rapport and trust with people that they relied on me to make sure that, to bridge it.

It is a truism that science is, and always has been, a collaborative endeavor reliant on the care of human relationships. But pressure, often from granting agencies, to collaborate more frequently and intensely across disciplinary boundaries has brought new complications. Collaborations centered on access to common data and data products can reveal sharp and surprising disciplinary differences and the precarious nature of scientific work threatens the continuity of scientific collaborations and careers.

Conclusion

Bush's annus mirabilis of 1945 cemented a sea change in the political economy of science. From the extraordinary productivity of wartime physics and chemistry, Bush envisioned a radically changed model of scientific inquiry and practice—changes cemented into American science with the creation of the National Science Foundation. A generation after Bush's annus mirabilis, Weinberg and Paisley observed that Bush's assumptions about this transformation were overly optimistic and failed to account for the vast archipelago of work and labor that propels and informs scientific practice.

Similarly, the introduction of data-intensive science has been accompanied by optimistic assumptions about the efficacy of making data digital and the ease of transdisciplinary research. Like Bush's vision of postwar science modeled after chemistry and physics, data-intensive science found strong affinities with a particular discipline, astronomy. Writing two decades from the dawn of data-intensive science, we, like Weinberg and Paisely, have examined and critiqued those assumptions. Our critique was grounded in an empirical analysis of the CKI's fifteen-year data corpus of continuous research into data-intensive science and guided by engagements with recent labor literature and classic readings in Science Studies. Throughout, we have drawn ethnographic examples from the extensive data corpus accumulated by several cohorts of researchers at the CKI. Though extensive, the CKI corpus and this article can only be a start in understanding the varieties and valences of labor in data-intensive science. Many consequential forms of (in)visible labor in data-intensive science have gone unaddressed by this article: the authorial and social work of securing funding under the grant system, the work of maintaining and repairing building and facilities, the labor of building precision instruments digitally connected to research infrastructure, and the labor of information technologists who maintain and repair the screens where much of data-intensive science takes place. Each of these areas encompasses a vast panorama of (in)visible work and labor that fall outside the purview of our data corpus.

The argument we have presented here assumes that a full and nuanced understanding of data-intensive science can only be obtained by starting with the in situ work and labor of scientific practice in all its manifold forms. We have begun but by no means exhausted this process. In particular, more research is required to understand how the process of seeking funding, and policy interventions, (such as open science and open data policies), bend and shape the contours of scientific practice and labor. As well, studies by CKI researchers have primarily concentrated on disciplines that have been early adopters of data-intensive techniques and practices. Yet today, techniques pioneered in those disciplines are spreading beyond the boundaries of the physical and biological sciences into the humanities, social sciences, and arts. Little is known about how the long-standing norms of work in those fields are being reconfigured as data-intensive techniques are introduced. More research taking the perspective of work and labor is needed to achieve a full and nuanced accounting of the changes brought where data-intensive techniques have taken hold.

Author Biography

Michael J. Scroggins is a postdoctoral researcher at the UCLA Center for Knowledge Infrastructures (CKI), where he works on issues of labor and competition in data-intensive science. Prior to joining the CKI, he completed a dissertation in Anthropology and Education at Teachers College, Columbia University, examining the organization and experimental program of DIYbiologists in Silicon Valley.

Author Biography

Irene V. Pasquetto is a scholar in the field of information and communication studies. She is currently a Postdoctoral Fellow at the Shorenstein Center on Media, Politics, and Public Policy at the Harvard Kennedy School (HKS), where she researches online disinformation. In Fall 2020, she is an incoming assistant professor at the School of Information at the University of Michigan. Before joining the Kennedy School, she was a research assistant at the UCLA Center for Knowledge Infrastructures (CKI), and a research fellow at the UCLA Institute for Society and Genetics.

Acknowledgements

Research reported here was supported in part by grants from the National Science Foundation (NSF) and the Alfred P. Sloan Foundation (Sloan): (1) The Center for Embedded Networked Sensing (CENS) was funded by NSF Cooperative Agreement #CCR-0120778; (2) CENSNet: An Architecture for Authentic Web-based Science Inquiry in Middle and High School, NSF #ESI-0352572; (3) Towards a Virtual Organization for Data Cyberinfrastructure, NSF #OCI-0750529; (4) Monitoring, Modeling & Memory: Dynamics of Data and Knowledge in Scientific Cyberinfrastructures, NSF #0827322; (5) If Data Sharing is the Answer, What is the Question?, Sloan #2015-14001.

Acknowledging Our (In)visible Labor

Funding provided by: The Alfred P. Sloan Foundation
Executive Producer: Christine Borgman
Lead Writer: Michael Scroggins (member, UAW Local 5810)
Assisting Writer: Irene Pasquetto
Data Curator: Milena Golshan
Copy Editor: Yookyung Bang
Developmental Editors: Bernie Boscoe, Morgan Wofford, and Milena Golshan of the Center for Knowledge Infrastructure at the University of California, Los Angeles. Peter Darch and Cheryl Thompson, University of Illinois at Urbana-Champaign. Special thanks to Tim Elfenbein and Shreeharsh Kelkar for comments on previous versions. Written in Los Angeles, California. Set dressing and production assistance provided by the University of California, Los Angeles. With special thanks to the staff of the GSE&IS for logistical support.

Bibliography

Allen, Davina. 2014. *The Invisible Work of Nurses: Hospitals, Organisation and Healthcare*. 1 edition. Abingdon, Oxon: Routledge.

Bell, Gordon, Anthony J. G. Hey, and Alexander S. Szalay. 2009. "Beyond the Data Deluge (Computer Science)." *Science* 323 (5919): 1297–98. https://doi.org/10.1126/science.1170411.

Borgman, Christine L., Peter T. Darch, Ashley E. Sands, Irene V. Pasquetto, Milena S. Golshan, Jillian C. Wallis, and Sharon Traweek. 2015. "Knowledge Infrastructures in Science: Data, Diversity, and Digital Libraries." *International Journal on Digital Libraries* 16 (3–4): 207–27. https://doi.org/10.1007/s00799-015-0157-z.

Bowker, Geoffrey C., and Susan Leigh Star. 1999. *Sorting Things Out: Classification and Its Consequences*. Inside Technology; Cambridge, Mass.: The MIT Press.

———. 2000. "Invisible Mediators of Action: Classification and the Ubiquity of Standards." *Mind, Culture, and Activity* 7 (1–2): 147–63. https://doi.org/10.1080/10749039.2000.9677652.

Burns, Randal, Joshua T. Vogelstein, and Alexander S. Szalay. 2014. "From Cosmos to Connectomes: The Evolution of Data-Intensive Science." *Neuron* 83 (6): 1249–52. https://doi.org/10.1016/j.neuron.2014.08.045.

Bush, Vannevar. 1945a. "Science -- The Endless Frontier: A Report to the President on a Program for Postwar Scientific Research (Reprinted 1990)." Washington, D.C.: National Science Foundation.

———. 1945b. "As We May Think." *Atlantic Monthly*, July 1945.

Collins, Harry, and Robert Evans. 2007. *Rethinking Expertise*. Chicago: University of Chicago Press. http://kutrov.web.elte.hu/tudhat/12_collins_evans.pdf.

Coulson, Margaret, Branka Maga, and Hilary Wainwright. 1975. "'The Housewife and Her Labour under Capitalism' - a Critique." *New Left Review; London* 0 (89): 59–71.

Cowan, Ruth Schwartz. 1976. "The 'Industrial Revolution' in the Home: Household Technology and Social Change in the 20th Century." *Technology and Culture* 17 (1): 1–23. https://doi.org/10.2307/3103251.

Crain, Marion, Winifred Poster, and Miriam Cherry. 2016. *Invisible Labor: Hidden Work in the Contemporary World*. Univ of California Press.

Critchlow, Terence, and Kerstin Kleese van Dam, eds. 2013. *Data-Intensive Science*. 1 edition. Boca Raton: Chapman and Hall/CRC.

Daniels, Arlene Kaplan. 1987. "Invisible Work." *Social Problems* 34 (5): 403–15. https://doi.org/10.2307/800538.

Darch, Peter T., Christine L. Borgman, Sharon Traweek, Rebekah L. Cummings, Jillian C. Wallis, and Ashley E. Sands. 2015. "What Lies beneath?: Knowledge Infrastructures in the Subseafloor Biosphere and Beyond." *International Journal on Digital Libraries* 16 (1): 61–77. https://doi.org/10.1007/s00799-015-0137-3.

Dominus, Susan. 2017. "When the Revolution Came for Amy Cuddy." *The New York Times*, October 18, 2017, sec. Magazine.

https://www.nytimes.com/2017/10/18/magazine/when-the-revolution-came-for-amy-cuddy.html.

Edgerton, David. 2017. "The Politcal Economy of Science." In *The Routledge Handbook of the Political Economy of Science*, 1st ed. Abingdon, Oxon ; New York, NY : Routledge, 2017.: Routledge. https://doi.org/10.4324/9781315685397.

Edwards, Paul N., Matthew S. Mayernik, Archer L. Batcheller, Geoffrey C. Bowker, and Christine L. Borgman. 2011. "Science Friction: Data, Metadata, and Collaboration." *Social Studies of Science* 41 (5): 667–90. https://doi.org/10.1177/0306312711413314.

Ekbia, Hamid, Michael Mattioli, Inna Kouper, G. Arave, Ali Ghazinejad, Timothy Bowman, Venkata Ratandeep Suri, Andrew Tsou, Scott Weingart, and Cassidy R. Sugimoto. 2015. "Big Data, Bigger Dilemmas: A Critical Review." *Journal of the Association for Information Science and Technology* 66 (8): 1523–45. https://doi.org/10.1002/asi.23294.

Fee, Terry. 1976. "Domestic Labor: An Analysis of Housework and Its Relation to the Production Process." *Review of Radical Political Economics* 8 (1): i–8. https://doi.org/10.1177/048661347600800101.

Foucault, Michel. 2012. *The Archaeology of Knowledge*. Knopf Doubleday Publishing Group.

Gray, Jim. 1996. "Evolution of Data Management." *Computer* 29 (10): 38–46.

Green, Toby. 2009. "We Need Publishing Standards for Datasets and Data Tables." OECD Publishing White Paper. Paris: Organisation for Economic Cooperation and Development. http://dx.doi.org/10.1787/603233448430.

Harding, Sandra. 2016. *Whose Science? Whose Knowledge?: Thinking from Women's Lives*. Cornell University Press.

Hey, Anthony J. G., Stewart Tansley, and Kristin Tolle, eds. 2009. "Jim Gray on EScience: A Transformed Scientific Method." In *The Fourth Paradigm: Data-Intensive Scientific Discovery*, xix–xxxiii. Redmond, WA: Microsoft Research.

Hills, Denise, Robert R. Downs, Ruth Duerr, Justin Goldstein, Mark Parsons, and Hampapuram Ramapriyan. 2015. "The Importance of Data Set Provenance for Science." *Eos* 96 (December). https://doi.org/10.1029/2015EO040557.

Himmelweit, Susan, and Simon Mohun. 1977. "Domestic Labour and Capital." *Cambridge Journal of Economics* 1 (1): 15–31.

Hochschild, Arlie Russell. 2012. *The Managed Heart: Commercialization of Human Feeling*. University of California Press.

Howison, James, and Julia Bullard. 2016. "Software in the Scientific Literature: Problems with Seeing, Finding, and Using Software Mentioned in the Biology Literature." *Journal of the Association for Information Science and Technology* 67 (9): 2137–55. https://doi.org/10.1002/asi.23538.

Jackson, Steven J. 2017. "Speed, Time, Infrastructure." *The Sociology of Speed: Digital, Organizational, and Social Temporalities*, 169.

Kitchin, Rob. 2014a. *The Data Revolution: Big Data, Open Data, Data Infrastructures & Their Consequences*. Los Angeles, California: SAGE Publications.

———. 2014b. "Big Data, New Epistemologies and Paradigm Shifts." *Big Data & Society* 1 (1): 2053951714528481. https://doi.org/10.1177/2053951714528481.

Lave, Rebecca, Philip Mirowski, and Samuel Randalls. 2010. "Introduction: STS and Neoliberal Science." *Social Studies of Science* 40 (5): 659–75. https://doi.org/10.1177/0306312710378549.

Leonelli, Sabina. 2010. "Packaging Small Facts for Re-Use: Databases in Model Organism Biology." In *How Well Do Facts Travel?*, edited by Peter Howlett and Mary S. Morgan, 325–48. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511762154.017.

Marcus, Adam, and Ivan Oransky. 2018. "The Data Thugs." *Science* 359 (6377): 730–32. https://doi.org/10.1126/science.359.6377.730.

Mauthner, Natasha S., Odette Parry, and Kathryn Backett-Milburn. 1998. "The Data Are out There, or Are They? Implications for Archiving and Revisiting Qualitative Data." *Sociology* 32 (4): 733–45.

Mayernik, Matthew S., Sarah Callaghan, Roland Leigh, Jonathan Tedds, and Steven Worley. 2015. "Peer Review of Datasets: When, Why, and How." *Bulletin of the American Meteorological Society* 96 (2): 191–201. https://doi.org/10.1175/BAMS-D-13-00083.1.

Mayernik, Matthew S., Jillian C. Wallis, and Christine L. Borgman. 2013. "Unearthing the Infrastructure: Humans and Sensors in Field-Based Scientific Research." *Computer Supported Cooperative Work (CSCW)* 22 (1): 65–101. https://doi.org/10.1007/s10606-012-9178-y.

Mirowski, Philip. 2011. *Science-Mart*. Harvard University Press.

Mol, Annemarie. 2008. *The Logic of Care : Health and the Problem of Patient Choice*. Routledge. https://doi.org/10.4324/9780203927076.

Moore, Allen J., Mark A. McPeek, Mark D. Rausher, Loren Rieseberg, and Michael C. Whitlock. 2010. "The Need for Archiving Data in Evolutionary Biology." *Journal of Evolutionary Biology* 23 (April): 659–60. https://doi.org/10.1111/j.1420-9101.2010.01937.x.

Morus, Iwan Rhys. 2016. "Invisible Technicians, Instrument-Makers and Artisans." In *A Companion to the History of Science*, 97–110. John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118620762.ch7.

Nardi, Bonnie A, and Yrjö Engeström. 1999. "A Web on the Wind: The Structure of Invisible Work." *Computer Supported Cooperative Work (CSCW)* 8 (1–2): 1–8.

National Institutes of Health. 2017. "Availability of Research Results: Publications, Intellectual Property Rights, and Sharing Research Resources." https://grants.nih.gov/grants/policy/nihgps/HTML5/section_8/8.2_availability_of_research_results_publications__intellectual_property_rights__and_sharing_research_resources.htm.

Orr, Julian E. 1996. *Talking about Machines: An Ethnography of a Modern Job*. Cornell University Press.

Paisley, William J. 1968. "As We May Think, Information Systems Do Not."

Pasquetto, Irene V., Christine L. Borgman, and Morgan F. Wofford. 2019. "Uses and Reuses of Scientific Data: The Data Creators' Advantage." *Harvard Data Science Review* 1 (2). https://doi.org/10.1162/99608f92.fc14bf2d.

Pasquetto, Irene V., Bernadette M. Randles, and Christine L. Borgman. 2017. "On the Reuse of Scientific Data." *Data Science Journal* 16 (March). https://doi.org/10.5334/dsj-2017-008.

Plantin, Jean-Christophe. 2018. "Data Cleaners for Pristine Datasets: Visibility and Invisibility of Data Processors in Social Science." *Science, Technology, & Human Values*, June, 0162243918781268. https://doi.org/10.1177/0162243918781268.

Ribes, David, and Geoffrey C. Bowker. 2008. "Organizing for Multidisciplinary Collaboration: The Case of the Geosciences Network." In *Scientific Collaboration on the Internet*, edited by Gary M. Olson, Ann S. Zimmerman, and Nathan Bos, 311–30. Cambridge, MA: MIT Press. http://www.davidribes.com/storage/RibesBowker_OrganizingforMultidisciplinaryCollaboration.pdf.

Roberts, Celia. 2018. "Practising Ambivalence: The Feminist Politics of Engaging with Technoscience." In *A Feminist Companion to the Posthumanities*, 199–210. Springer, Cham. https://doi.org/10.1007/978-3-319-62140-1_17.

Russell, Andrew L., and Lee Vinsel. 2018. "After Innovation, Turn to Maintenance." *Technology and Culture* 59 (1): 1–25. https://doi.org/10.1353/tech.2018.0004.

Scroggins, Michael. 2017. "Ignoring Ignorance: Notes on Pedagogical Relationships in Citizen Science." *Engaging Science, Technology, and Society* 3 (0): 206–23. https://doi.org/10.17351/ests2017.54.

———. 2019. "Designing, Animating, and Repairing a Suitable Do-It-Yourself Biology Lab." In *Educating in Life: Educational Theory and the Emergence of New Normals*. Routledge.

Shapin, Steven. 1989. "The Invisible Technician." *American Scientist* 77 (6): 554–63.

Souleles, Daniel, and Michael Scroggins. 2017. "The Meanings of Production(s): Showbiz and Deep Plays in Finance and DIYbiology." *Economy and Society* 46 (1): 82–102. https://doi.org/10.1080/03085147.2017.1311134.

Star, Susan Leigh. 1990. "Power, Technology and the Phenomenology of Conventions: On Being Allergic to Onions." *The Sociological Review* 38 (1_suppl): 26–56. https://doi.org/10.1111/j.1467-954X.1990.tb03347.x.

Star, Susan Leigh, and Anselm Strauss. 1999. "Layers of Silence, Arenas of Voice: The Ecology of Visible and Invisible Work." *Computer Supported Cooperative Work (CSCW)* 8 (1–2): 9–30. https://doi.org/10.1023/A:1008651105359.

Tronto, Joan. 1993. *Moral Boundaries: A Political Argument for an Ethics of Care*. New York: Routledge.

Tyfield, David. 2013. "Transition to Science 2.0: 'Remoralizing' the Economy of Science." *Spontaneous Generations: A Journal for the History and Philosophy of Science* 7 (1): 29–48. https://doi.org/10.4245/sponge.v7i1.19664.

Velden, Theresa, Matthew J. Bietz, E. Ilana Diamant, James D. Herbsleb, James Howison, David Ribes, and Stephanie B. Steinhardt. 2014. "Sharing, Re-Use and Circulation of Resources in Cooperative Scientific Work." In *Proceedings of the Companion Publication of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing*, 347–350. CSCW Companion '14. New York, NY, USA: ACM. https://doi.org/10.1145/2556420.2558853.

Wallis, Jillian C., Elizabeth Rolando, and Christine L. Borgman. 2013. "If We Share Data, Will Anyone Use Them? Data Sharing and Reuse in the Long Tail of Science and Technology." *PLoS ONE* 8 (7): e67332. https://doi.org/10.1371/journal.pone.0067332.

Weinberg, Alvin M. 1961. "Impact of Large-Scale Science on the United States." *Science*, New Series, 134 (3473): 161–64. https://doi.org/10.1126/science.134.3473.161.