Making & Doing

Activating STS through Knowledge Expression and Travel

Edited by: Gary Lee Downey, Teun Zuiderent-Jerak

Citation:

Making & Doing: Activating STS through Knowledge Expression and Travel
Edited by: Gary Lee Downey, Teun Zuiderent-Jerak
DOI: 10.7551/mitpress/11310.001.0001
ISBN (electronic): 9780262366052
Publisher: The MIT Press
Published: 2021

OA Funding Provided By:

The open access edition of this book was made possible by generous funding from Arcadia—a charitable fund of Lisbet Rausing and Peter Baldwin.

The MIT Press

2

THE GROUND KEEPS OPENING UP

Building an Infrastructure for Data Appropriation

Dawn Nafus

THE FIRST MEETING

Finally, we were really getting started. After much research and concept development, we were finally in the nuts and bolts of prototyping Data Sense. The prototype was a web-based data exploration software for people who use data to keep track of a health or personal problem—self-trackers. Self-trackers do not necessarily have the same skills as a data scientist but often ask questions about what data could mean beyond the meanings that are already supplied by tracking devices. Normally my job consists of research to identify where new technology or business opportunities lie, and this was the first time I would be involved in a hands-on way. I wanted to start working hands on because helping people work with data on their own terms mattered to me. Asking questions requires interpretive work beyond consumption. The interpretive work is not just statistical; it involves critical reflection about the circumstances in which the data was collected. Because this was both intellectually interesting to me and, in my view, socially beneficial, I cared deeply about whether it was possible to support self-trackers in this process, and I was lucky enough to be working with collaborators who similarly saw the value of doing so with software.

I work in a research arm of a large company, Intel, where the connection between research and product development is never straightforward. Markets for technology coalesce in complex ways that are never visible at the outset, which creates an unsteadiness that can be disorienting and alienating, not at all like the formerly celebrated and now panned out-of-the-box, disruptive thinking Silicon Valley became known for. If we were crazy to pursue this, it was a thoroughly researched, highly experienced form of crazy. In their work as engineers, or in their personal lives, my colleagues on the team knew the value of data exploration as the first step to meaning making. They also knew that context was the key to establishing data's meaning, so it made sense to them that my ethnographic research had shown that self-trackers were good at contextualizing their data and did not always take the messages from tracking devices at face value. In fact, one self-tracker in our early research was so careful with his data, both mathematically and conceptually, that one of the computer scientists on the team exclaimed, “This guy’s my hero!” Indeed, others on the development team were themselves self-trackers. These confirmations from multiple perspectives softened my anticipation of being told I was crazy yet again, which had come to take on a more stinging valence over the years, being both a woman and a nonengineer in a male-dominated tech company.

In these circumstances, I felt a deep sense of responsibility to get it right. This first daylong team meeting turned to discussion of the types of data that would be ingested into the system—strings of text? Numbers? URLs? Enum (or answers to multiple-choice questions)? Something called a float? I didn’t understand why we needed to anticipate all these things beforehand, but I took it on trust that we did and secretly wondered to myself what it could possibly mean for a number to float. It became important to know whether time series data would be the only kind of data that we would need to bring into this system. “Dawn, do self-trackers collect any data that isn’t time series?” I had no idea. I could tell them all about the politics of governmentality and measurement, but whether only time series data was in play, I had no idea. I could guess that time series was probably the most common data type—keeping track involved a notion of time, after all—but whether other data types were of note was a mystery. This prompted a debate about what we meant by data “types”—enum and float or heart rate and blood pressure?—and how much definition we needed at this early stage.

At some point, our newly hired full-stack[1] developer said that we didn’t have to solve all the corner cases just yet. I had to ask about this one. If I didn’t understand the design and development process from the beginning, there was no way I’d be of much use. “What’s a corner case?” I asked. Quizzical looks turned my way. It turns out that defining a corner case was something I was expected to be able to contribute to. It is a scenario of use that is rare but could be necessary to take into account when designing a system. Surely it was my job to define what those are, so exposing myself as not even having the language for it would have earned me the label of “crazy” in a less forgiving team. Later I would come to understand this forgiveness not only as a sign that I was finally working in an inclusive team but also as a sign that we were not building a data visualization tool, but rather an infrastructure where the unknowns would proliferate and put everyone on equally unsteady footing. Exploratory data analysis required us to contemplate design considerations that ranged from database issues (in tech industry speak, low in the stack[2]) all the way through to considerations that involved no technology at all, up so high it was above the stack entirely. Exploratory data analysis complicates any naive understanding of what serves as infrastructure. It is a sociotechnical process that makes it impossible to assume which elements constitute the visible figure and which the invisible ground.

This chapter describes what is involved technically and socially to support that kind of analysis and how my own assumptions about data and infrastructure unraveled for me. Data is regularly appropriated or contested by supposed nonexpert groups like self-trackers and air quality activists. Yet supporting alternative interpretations of computationally dense data sets requires more than alternative social and cultural frameworks. It also requires alternative technical infrastructures to support the formulation of new stories that are told with data. Our infrastructure exposed yawning gaps between computational systems and social worlds, which manifested as uncertainties or unknowns in the design process, despite decades of datafication (Van Dijck 2014). Our design ethos was to help people *find* the story in data rather than telling the story to them or for them. This ethos opened up a whole world of unknowns, both technical and social, precisely because the work of finding the story in data is unpredictable. The unfolding of unknowns exposed a double invisibility: it made the invisibility of infrastructure become visible, which in turn exposed the equally invisible nature of exploratory data analysis work, which is not about making bold claims and storytelling but about doing the relatively unglorious work of merging data sets, shipping them between systems, and finding roads in the data that don't reveal much meaning. By building a data infrastructure, we found that in contexts in which data work lacks a social location—a defined occasion for data wrangling and merging to happen—the appropriation of data is deeply constrained. We also found that the unfolding of unknowns can be experienced viscerally, sometimes in the form of excitement but often in the form of worry.

To tell this story, I trace the project in rough chronological order, from the earliest research through development, testing, and its somewhat ambiguous endpoint, showing what opened up at each step along the way.

THE WORK OF SELF-TRACKING DATA

Data Sense started life in a research program known as Biosensors in Everyday Life.[3] This was a three-year, four-university interdisciplinary research program organized and sponsored by Intel, where I served as the program manager. It was a program that set out to better comprehend the social dynamics that product planners would encounter should they choose to invest in consumer-grade biosensors, which are a type of sensing technology for detecting substances in the body or in the environment. It is useful to know whether the social dynamics would support or inhibit markets for such devices and what the likely usage scenarios were. In *Quantified: Biosensing Technologies in Everyday Life* (Nafus 2016), the investigators from that program and I described in depth the sociological and anthropological questions addressed in that research. What might it mean for medical knowledge and institutions to have consumers appropriating medical technologies? What are the consequences of an expanded role for data in daily living? How should we think about notions of public and private when multiple people are implicated in the same data set and when a data set is nominally about an individual but in reality about so much more? What are people actually doing with biosensors and their predecessors?

These were large, complex problems in their own right, and it was my job to "distill" all of this research for those who were making business decisions with it (an example of the distillation can be found in figure 2.1). The quotation marks indicate that my translations were hardly a distillation of the essence of the research; they were an opinion about what was important. My academic colleagues on that program indulged me in an exercise in which, for each social dynamic they identified, we made some structured speculation about how liable to change each was. For example, if there was a widening gray area between medical knowledge and lay knowledge, would this expand or contract in the coming years? What could the consequences be? Through this structured speculation, the answer I ended up giving to the business about whether a market was feasible was short and blunt: there were lots of enabling dynamics and some inhibiting ones, but the inhibitors would not render the market unviable. I said that privacy was going to be incredibly difficult to get right and would require real investment to do so.

The research program informed a number of business decisions beyond Data Sense, but here I focus on how we arrived at understanding there was a need for software of this kind. It started in the strand of the research focused on data literacy. At

Questions Intel asked in 2010	What we learned
• Is the health care industry going to adopt consumer biosensors as an extension of care?	• Not easily or quickly. Consumer data makes work and liability for doctors. Data that matters to consumers is not what matters to doctors.
• What usages make sense?	• Consumer-driven biosensor data is extremely helpful when just getting a pill doesn't work (allergies, asthuma, fatigue, migraines, etc.).
	• Domains with a lot of cultural baggage (weight management, genetics, etc.) find it harder to deliver real value, because the necessary action is out of the user's control.
• Which sensors have the greatest potential?	• Patterns and triggers are heterogeneous. Next-generation capabilities matter, but no one sensor is the "killer app."
	• Combining environmental sensing and personal sensing currently is ripe for innovation.
• How do we present scientific data in a simple way?	• People can (and do) make sense of complex datasets when it matters to them. Sheer complexity isn't the problem, putting it in context is.

1

2.1 A transformation of scholarship into technology company frames. This slide was used at the final meeting between Intel stakeholders and Biosensing in Everyday Life research program investigators.

the time, my strategic planning colleagues had a concern about whether health data could be made simple enough for nonspecialists to interpret in an appropriate way. There was, of course, a heavy dose of medical paternalism in the framing of the question. The fears in the medical community about people becoming unduly frightened about scary-looking data are wildly out of proportion to the rate at which this actually happens (Brown 2016). This medical aspect was compounded by long-standing tropes within the technology industry regarding the ill-informed computer user, who is the imagined source of error in otherwise well-designed programs (Woolgar 1990). "Prosumers," or lead users, who do technical work in the act of consumption are acknowledged, but fantasies of machinic control—that the device would nudge the consumer to act in certain ways—loom large. As wearables emerged as a product category and intersected with medical notions of compliance, the figure of the prosumer had become eclipsed by a quasi-helpless consumer figure who required guidance via automatic notification to stop or start a certain behavior (Dow-Schüll 2016). The work she needed to do was to discipline herself in response to the numbers that devices delivered, such as taking ten thousand steps. She was not expected or trusted to do the intellectual work of interpreting numbers. To her, they were supposed to be both self-evident and self-efficacious.

However, this strand of research showed that there was substantial room for alternatives. We found that there was no such thing as simple or complicated data parsings or calculations; what mattered was what people were attuned to. This work built on Lave (1988) and Verran (2001), who point out that the concept of multiplication is not inherently more complex than the concept of length and that the ability to get it has to do with whether the mathematical procedure is in tune with ready-to-hand cultural and linguistic structures. For example, Euro-Americans know what a percentile is not because we know it mathematically but because most of us have been measured in those terms. Building on these ideas, investigators from Goldsmiths College (Day, Lury, and Wakeford 2014) showed that this cultural process of attunement continues in the context of big data systems. In encounters with social media metrics, for example, lay people show a tremendous facility with the sorts of calculations made possible only by large-scale, complex data sets (Gerlitz and Lury 2014). Similarly, a team from Lancaster University discovered the extraordinary lengths some people would go to in order to do data mining on their own direct-to-consumer genomics test results (Kragh-Furbo et al. 2016).

My own parallel research on self-trackers in the Quantified Self community (Nafus and Sherman 2014) showed that the willingness to spend time with data and to derive meaning from it was not necessarily a function of having STEM (science, technology, engineering, and mathematics) skills but more a function of being in a situation that required good data, such as health conditions like insomnia or migraines that could have an avoidable trigger in everyday living. People with these conditions are often willing to spend time trying to understand patterns because they could very well discover ways to prevent their symptoms. In some ways, they also stand a better chance of success than a professional statistician or clinician if the issue is better characterized with data from outside the clinic than inside it. Through this research program, the researchers and I shifted the problem space away from questions of how to make data simple enough to defy any reinterpretation and toward questions of how to design for appropriation. That is, how could we support users' interpretation of sensor data as a function of context?

At this point we had an intuition that current tools were not doing this job well. For example, one self-tracker in my research had combined a daily step count with autoimmune symptoms data and learned that, over the long term, taking too much exercise was triggering her autoimmune disease (Wheelwright 2016). She made this discovery in part because she happened to use a fitness tracker that offered aggregations of steps by months, not just days. Most steps counters offer daily step counts because daily management of exercise is the manufacturer's intended use, but daily aggregation would have made this long-term pattern visually imperceptible. If her problem had been related to a seasonal recurrence, a different graph again might have been needed. If she had three years' worth of steps data, to create such a graph would require advanced knowledge of the right functions in Excel, if indeed the data can be put into Excel at all. If it had been recorded once a minute, most spreadsheet programs would stop working because the file size starts to get too big after three years.

Seeing the pattern clearly also requires the ability to annotate the graph with a timeline of symptoms, which is a capacity largely unavailable on sensing devices or visualization tools. And what about spatial patterns or patterns related to another source of data, like weather or air quality? The latter would require knowing how to deal with time stamp formatting and time zone alignment, which is a problem that people don't usually encounter unless they are developing software. The goal in this context, though, is not to develop software but to get to the point where it is possible to identify spatial or temporal recurrences. In this case, Jacqueline Wheelwright was able to pull together the right data skills and interpretive skills to find meaning in her data that was useful to her, but there are many other circumstances in which she very well might have hit a wall and not found meaning.

If we take a step back, then, and consider the figures for whom the relevant tools have been designed, we can see a split between tools for general-purpose data wrangling, in which the point is to maximize the available ways to parse data, and tools for nontechnical people—consumers, not prosumers—that minimize the possible parsings available and communicate only the designer's intended meaning. Clearly, this divide stems from market assumptions about who does what kind of data work. That split, however, does not hold up to ethnographic reality: there are indeed people who are not engineers or data workers but who are prepared to do some exploratory data work. Such people make for an ill-defined cultural figure, however. Only nascent cultural and social formations support them, like the Quantified Self community,

community-based participatory health and environment research, and now, the MyData movement (see Lehtiniemi and Ruckenstein 2018). Data exploration is largely invisible work that a person does, not who a person is. This is true for both people who do not identify as geeky and for people in professional data science circles, where exploration and cleaning is widely acknowledged to be 90% of data work, yet this is rarely taught explicitly (Dumit 2018). Data exploration does not contribute to identity formation in the way that learning to code and learning data science feature in the cultural imagination.

This third figure, neither coder nor consumer, can come from either those who see a problem and cobble together tools or ramp up on the data skills needed to understand it or those who already have data skills and learn to do a deeper, more critical reflection of the data's context (for a description of those skills, see Loukissas 2019). In practice, the third figure comes more often from the latter group than the former. This is in part because of a lack of cultural location for this third figure—getting to the point of needing data can be a sign of the direness of one's situation, not identities or ambitions. It also stems from the high level of investment required to deal with conceptually irrelevant matters like time stamp alignment. Even professional data wranglers spend an enormous amount of time dealing with such issues, and what constitutes a lot of time for a professional will stop the works entirely for someone else. Therefore, the Data Sense team saw value in automating what could be automated and accommodating file sizes that break spreadsheet programs. The alternative is either working at a command prompt, which requires an interest in programming per se that few self-trackers have, or working with a data visualization program, which contains useful visual tools but few computational tools like aggregation functions or time series correlations.

The gap between professional data wranglers and self-trackers can also be seen in the practice of churning through visualizations before finding the one that speaks to something meaningful. Self-trackers with data skills told us pointed stories about converting data that recorded the days when a certain activity happened to the days when it was skipped. A person needs to have faith that the conversion would show something at all in order to go through all the pain to render it visually! Such faith is not well founded, though. Usually, finding meaning in data means tacking back and forth between the parsing and the visualization, ending in multiple parsings and visualizations that prove meaningless. It means toggling between days of an activity and the days skipped or churning through different data sets brought in and out of the same graph. This tacking back and forth between computation and graph, recalculation and new graph, and then back to the old graph for comparison is central to exploratory data analysis, whether doing self-tracking or in the lead up to writing an algorithm. However, what was clunky, slow work for computer scientists was impossible for others. In fact, one data visualization professional recommended to me, as someone learning data skills, that the best way to find the right visualization is to start by sketching and avoid the pain of doing it computationally until I was more confident of the right direction.

These constraints in the work process showed us that very specific technical support was required before this third figure stood a chance of reframing the meaning of data. Reducing the programming work required for data exploration was a problem that we could make headway on as an engineering lab. It would be an experiment, but what kind of experiment? It was a commercial experiment in the sense of having a long-term commercial goal of enabling user-driven innovation in a market that clearly needed it; our company stood to benefit from overall growth as the supplier of server-side and client-side hardware. This commercial motivation meant that we focused on digital data rather than analog data appropriation (which has its own enormous advantages—see Guggenheim, Kräftner, and Kröll 2018). However, it was not a commercial experiment in the sense that there was no attached business model or marketing that was also undergoing testing.

It was also not entirely a social experiment in the inventive methods sense, in that it was not designed to optimize for social research gains (see Marres, Guggenheim, and Wilkie 2018). Still, it was related to that approach in that we were radically open to all sorts of possible social outcomes, and we were learning something new about data cultures. Our positionality informed our choice to focus on those who believed they needed data but did not have proper access to it (Kennedy and Moss 2015), as opposed to those who had too much data and did not need it but nevertheless suffered the repercussions of others possessing data about them (Van Dijck 2014). Beyond that choice about audience, we were not committed to any preconceptions of what self-trackers would do with such a system, if anything, although obviously, one cannot build a tool without hoping someone will use it in some capacity. Which capacity, however, was not something we were attached to.

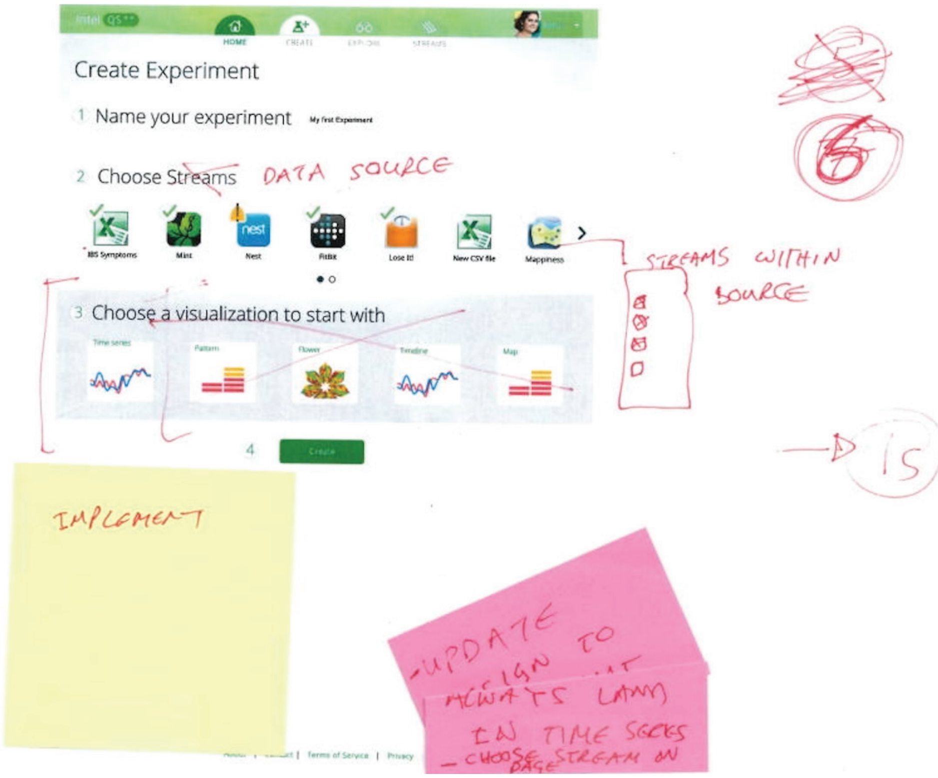
It was a quasi-experiment in the engineering sense, too. To explore what it takes to facilitate data appropriation, a working system was necessary. It need not be full production quality, but it did need to go well beyond a nonfunctional demonstration, or a set of tests of specific components, or a "hacky" effort to pull together off-the-shelf features.[4] It was a matter of engineering research to identify the technical approaches necessary to handle data of these types and sizes, just as much as it was a matter of social scientific research to explore how people might encounter data in circumstances in which some programming tasks were automated. Thus, it was not a product development exercise, but it still required building an operational, full-stack system that looked almost like a minimum viable product.[5]

FIGURE AND GROUND

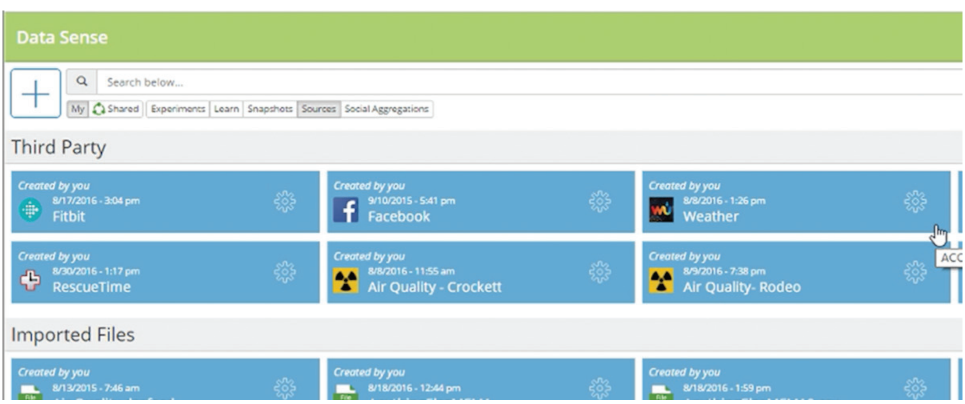
Exploratory data analysis, in which the analyst tacks back and forth between visualization and calculation, forced us elaborate the connections between figure and ground. In working on Data Sense, I found myself relearning in an embodied way the invisible nature of infrastructure as described by Star (1999). Infrastructure became very visible to us, not because it broke down, but because it had to be built up in the first place. We knew that in order to render data visualizations at all, we would have to offer some means of account log-in, data ingestion, and storage and management. That is, users needed to have a way of knowing what data exists in their account and to have a means of pulling it in and out of the system and in and out of a visualization within the system. These things were the ground to our figure of exploratory visualizations, which in turn had to be malleable enough, and traceable enough, to serve as ground for the larger process of making meaning. We contracted with a local design firm in order to wireframe (a middle stage between concept design and full-fledged design specification) the initial version of the website. That work supposedly included both ground and intermediary figure of visualization. We quickly discovered that we were barely able to get through the "ground" part in the time we had allowed for both.

The more areas of the ground we defined, the more they demanded specificity, which opened up yet more ground. Even a wireframe design could not proceed without knowing how many data streams (such as heart rate and steps) a typical sensing device offers. If we wanted to facilitate importing data from a service that had five data streams, we might need to design a simple drop-down menu, but at twenty data streams a different interaction would be necessary (see figures 2.2 and 2.3). This forced us to confront the issue that people might think of their data only in terms of the source, like "my Fitbit data," and not as a set of individual measurements, and so we needed to devise ways to preserve the ability to identify the source alongside the individual referent. Would users want to create derivative data sets, and if they did, were the data sets interstitial and throwaway, or a final product to be stored separately? Would people care about the ability to constantly update their work with live data feeds, or did they see their data as already collected? Would updating be affected by the presence of derivative data sets, which are created at one moment during analysis? In building databases, back-end software systems, and front-end interfaces, we need to make an assumption about these things, yet we had none because the tool had not yet been built.

Questions of how many sources people would have, how they would organize them, and how often they would need to rename the sources or the streams needed specification before we could make appropriate interface and database design decisions. These were interstitial features of the ground of exploratory data analysis, but they kept pushing into the figure. These types of questions were really infrastructural questions, and that is the reason I found myself in a discussion about data types, and not visualization, in the team's first meeting. The infrastructure became visible upon building it and an area of focus that necessarily situated the user as background. That is, we needed to know what manipulations were implied when a self-tracker says she



2.2 Hand-drawn markup of our initial concept design when we realized that structuring data access was more complex than initially thought. These changes suggest a different way of handling data to reflect the difference between a data source, like "my Fitbit data," and a data stream, like "heart rate."



2.3 Final design of the live system. The feed-like structure accommodates any number of items of any type. We changed to a feed structure after having discovered that users have significantly more data than originally anticipated. Unlike drop-down lists, feeds accommodate an infinite number of items.

tracks the days she skipped something (or whatever she did). How they coalesced into a story was hard to keep in sight with so many underlying questions pressing.

As in a cascade, defining how data moved in and out of the system only opened up yet more ground. The question "How much data do people have?" proliferated into ever more specific aspects and technical requirements. System testing showed just how much the sheer number of megabytes affects the ways of bringing data into the system. Smaller files (i.e., data sources) could load within the time it takes for the browser to time out; larger ones needed background processing. Similarly, when trying to render the data on a screen, how many bits each individual data stream contained affected what we could render. If a data set has a million points, trying to represent them all would lead to the browser timing out before any line was shown at all. The original algorithm we used to render a line graph picked points at regular intervals (say, every fifth or twentieth data point, depending on the size of the data set). This led to spikes or dips being left invisible and users believing their data had gone missing. In fact, it is so easy to think of an algorithm as something that makes a classification, and a line graph as something that contains raw data, that when the error first happened to me I felt a visceral unease in my stomach, so disorienting was it to see my raw data—my grounds to an analysis—being reduced in a way that was not faithful to the spikes and dips that I had come to believe characterized the data. No amount of believing the STS exhortation about raw data being an oxymoron (Bowker 2005) can remove that bodily disorientation. As with other infrastructures, none of these things matter to the intended use of data storytelling per se, but became crucial bottlenecks when we got it wrong.

Once the grounds of data ingestion and management were mostly settled, we found ourselves elaborating an entirely new category of ground around the visualizations. We knew that self-trackers did not go directly from raw data to the exact visualization that they could make meaning from. We also knew that self-trackers tended to have on hand more data than is reasonable to look at in the same visual field. A Fitbit alone yields twenty-seven separate data streams. For these two reasons, we needed to support interstitial activity that would allow people to go back and forth between heterogeneous data streams. We arrived at the notion of an experiment in which users could hold steady a small handful of data to visualize in multiple ways, without having to load data and then render it for each new visualization. One account could have many experiments loaded with different configurations of data, and the state of each experiment would be saved automatically, much as Google's online office tools do. This provided a scaffolding to keep many iterations going in parallel, which we thought might better reflect how people do exploratory data work.

However, other tools had no equivalent of our experiment feature, so we had no idea how many experiments users would want in their account. Two? Five? A deep list of them? We initially underestimated the number of data sources and streams people would have (whether it was Fitbit data plus a file, or twenty files with ten parameters each, or something else) and the number of experiments they would want to keep running at the same time. This led to a costly and lengthy redesign of the ground, right in the middle of the project. Although this was hugely frustrating for everyone involved, our initial wrong guesses said something about the invisibility of exploration work. It showed palpably, and indeed painfully, that there are more data points and iterations in exploratory data analysis than one might assume, even if you have observed self-trackers' practices closely, and even if you have done exploratory data analysis as a professional. It exposed a forgetting of sorts built in to data exploration practices. All the visualizations that didn't work, and those other data sets that didn't turn out to matter, fade away from memory. The forgetting made it difficult to estimate the number of roads traveled in infrastructure building.

That experienced software developers, designers, and engineers shared my hope that we could get to the "figure" part (the visualizations) relatively quickly speaks to an even deeper sense in which infrastructure is inherently invisible. I was the only one new to the design and development process, yet we all shared the hope that we could begin designing the "visualization" part in the same round of work as the "infrastructure" part. I developed the suspicion that perhaps all this talk of agile development methods and iteration was really a way for developers to deal with the inherent invisibility of infrastructure. Although it is easy to deride the tech company ethos of agile methods, and its more nefarious twin of "move fast and break things," I came to understand that once we cast ideologies of speed aside, there is an underlying recognition in these metaphors that one cannot know what one does not know without having done it yet. The only way to learn is by attempting to concretely define in detail a working system and expose more unknowns. It is no fun to see a chasm open up of yet more ground to be developed, which in turn requires yet more budget negotiations to get the development done, and I did indeed find a coping mechanism necessary. The notion that this was a known problem, an agile process, seemed to help. A better description of "agile" might be "move slow and learn things" because new grounds kept opening, and yet they still remained grounds. Even fast production seemed slow at the surface. This transformed the notion of invisible infrastructures from a social scientific analytic framework to a sense of vertigo, felt bodily in the anticipation of yet more work that lay ahead.

HOW MANY AND WHAT KINDS OF FIGURES?

Once we did have a working system, or at least a skeleton of a working system, the process of tacking back and forth between visual forms—and forms of calculation—and interpretation, came to the fore. That is, the visualizations that were necessarily treated as "figure" in the building process became "ground" in an interpretive process. We knew, for example, that no single visualization would tell a story in the way data visualization gurus like Edward Tufte (1983) evangelize. A visual story is a product of refinement, a claim about the story worth telling, whereas our visualization features were visual tools to *find* the story in the first place. What would make a good visualization in our tool would be forms that get users part of the way there or help them eliminate wrong explanations or dead ends as much as find them. To identify what these visual forms should be, we held a series of Quantified Self community events where a small group shared data with one another, and we participated by making some guesses about ways to process it. It turned out there were some clear patterns to the parsings that participants found to be valuable, such as the time of day when something recurs or the ability to smooth out data that has lots of variability to see an underlying trend. In this way we made reasonable guesses, but ultimately one can only ever guess if the story itself is an emergent property of the interaction between person and system.

If the point is to open ground in which stories stand a chance of emerging, the temptation to add feature upon feature is pressing. This is a normal temptation in product design, but it is complicated here by the acknowledgment that a story could emerge from data in a multiplicity of ways. Heated discussions in our group occasionally ensued about what features were worth adding. Adding control over numerical parsings—the ability to change whether a recurrence by time of day was aggregated by averaging or summing points, for example—was a very different thing from adding visual controls, like control over the color of lines or bars. Our interaction designer unsurprisingly advocated for more visual control because that is where he found meaning. Some of the engineers on the project pointed out that without those numerical controls, the visualizations might not make any sense. Data about steps ought to be summed to get steps per hour or day, but adding up heart beats across an hour or day has no meaning. Indeed, so wildly heterogeneous is our current data culture that our users were not in agreement about which features were easy to use and intuitive and which were hard.

The dimensions of data stories proliferate as different data sets are combined. One datum is unlikely to tell a story, but data in a time series might have a story whereas many data sets across that same time frame might reflect different aspects of the same phenomenon. So the temptation to add one more data type or one more computational or visual control grew stronger the more we tested it with people because they often suggested that the stories they wanted to be able to tell would be possible if only Data Sense could do just that one extra thing. The challenge was ensuring enough of a range that someone could get to a personal discovery of some kind but one not so extensive that cognitive overload sets in because of the need to learn the interaction mechanisms and what those mechanisms do to the underlying data. The computations that the interactions performed are not always trivial to explain. To go back to the earlier example of Wheelwright, the woman who identified her autoimmune triggers in her steps data, she might not have recognized that an aggregation function, buried in the advanced-user corners of the interface, was relevant to her, even though changing the aggregation from average steps per day to total steps per month is effectively what she did.

The difficulty of deciding which visual figures would be good grounds for finding a story and of figuring out the best way to show the options to end users without overwhelming them revealed a misalignment of social worlds. In the very act of trying to bring together humanistic, context-sensitive interpretation with contemporary computational capabilities, moments like these made the gaps between those two worlds seem bigger than ever. Fellow data wranglers, and developers working on similar systems for those who *could* code, easily recognized what we accomplished. They saw our work as making a giant step toward easing the work of nonspecialists. Yet those nonspecialists were still seriously challenged, at least some of them. Recognizing both realities is dizzying, almost as crazy making as software makers recognizing both realities was dizzying. Even though we had seen plenty of examples of people who had defied stereotypes of what a nontechnical person can do, we came to understand that our attempt at developing a computational infrastructure around what they had accomplished did not hint at a new market but instead exposed a tremendously complex and subtle set of skills that some self-trackers had built for themselves. There were no simple ways of tidying this problem. There was no avoiding, for example, the need to add controls over the aggregation function (e.g., taking the average vs. summing) that are used to find recurrences by hour of day or day of week. Similarly, if we were going to show correlations, we would need to explain how a correlation coefficient worked and why time series correlations break down in certain circumstances. These were functions we could not make invisible because, in going down different roads to find the story, people would have to encounter them. We could not make the entire world of computation invisible in the way that we can make time stamp alignment somewhat invisible.

As soon as we started to see these limitations crop up in testing the system, we took steps to overcome them by thinking carefully about onboarding. Onboarding involves designing system elements that introduce new users to its mechanisms. For us, this meant introducing users not only to the interface but also to computational concepts. We were careful to not assume prior statistical knowledge, and instead we explained our terms. In fact, the exercise of explicitly naming all the capabilities, tooltip by tooltip, video tutorial by video tutorial, was our way of reinscribing, or otherwise shoring up, this nascent cultural location for data work that was barely emerging. Our onboarding features were radically different from introductory statistics instruction in that they focused on manipulating time series data to find patterns, not probability or statistical significance. This is a different genre of thinking, or dialect, of statistics. It was in part rooted in our sense of data as a "third thing" (Dumit 2018, 271), neither strictly quantitative nor qualitative. Here, numerical data triggered a qualitative analysis of a context because the viewer would know what else was going on in her life and why data might spike on a certain day of the week. That viewer does not always need a test of statistical significance to determine whether it mattered. This data dialect is one that few people speak all the time, and, as with a dialect, most are better at hearing and understanding than speaking. It would be perfectly interpretable to present a self-tracking finding in this dialect at a Quantified Self meeting. One could show, for example, a map that marked the average calories consumed in this location versus that. In fact, we built a mapping feature because we had seen some similar visualizations at Quantified Self. However, the conceptual frameworks necessary for a new Data Sense user to *make* that same map were not yet naturalized in the same way that a notion of an average has been naturalized enough to interpret the map. A user would need to know how to import the data, change the aggregation function, and then figure out that location data not only referred to where the user was but also was associated with other data in the account if it overlapped in time. The user would have to know not to expect "my location data" to show up as pins on a map, but that location data would be used as context to spatially analyze other data. A user would also then have to have an intuition that a spatial pattern of calorie consumption could be worth exploring, but if it looked odd on that map, it would be worth cleaning out spurious zeros. Even though we had done a tremendous amount of programming labor to make users not have to think about how to associate location with data or not deal with the zoom functions that aggregate and disaggregate data at various spatial resolutions, making that map is still fundamentally not easy because this is not an analysis most of us have to do on a daily basis. The need to design onboarding features points to a paradox, one in which we were actively participating in the emergence of a cultural location[6] by naming useful parsings of data, yet the need to name them at all reminds us just how emergent this location is and just how little one can assume about what others know.

CREATING A CULTURAL LOCATION, FOR A WHILE

User tests and community-based participatory research served as appropriate and useful cultural locations where the practices of this third figure could emerge. I was responsible for the ongoing user testing for the project, and we tested a range of possible end users, from those with data skills looking for some level of automation to those who saw a need to manipulate data but had no notion of where to begin. For the latter group, I often started an interview by taking the lead and doing a walk-through of the person's data using the tools. Participants absolutely were capable of both bringing their interpretive capacities to bear and learning about new aspects of data they hadn't considered.

For example, I did extensive user testing with a grassroots community group interested in examining the effects of air quality on health. Cardiovascular data like heart rate and blood oxygen from consumer-grade wearables and data from a local continuous air quality sensor were compatible enough to correlate. The data set sizes were right for Data Sense because Data Sense could handle amounts of data that spreadsheets cannot but doesn't require commercial-scale big data infrastructures. Our tool and their needs seemed like a good fit, and we worked out a joint experiment in which community members would collect cardiovascular data on themselves, and I would support their analyses of the data using this newly made tool. This in turn enabled me to test whether the tool's features were working and how users interpreted them.

This community was very sophisticated about the dangers of data smoothing because air quality regulations were written with respect to time-based averages. A polluter would be out of compliance if it had emitted benzene above a certain threshold on average for twenty-four hours or for a year. Community members complained that water can be 70 degrees Fahrenheit on average for twenty-four hours, but a burst of boiling water within that time can burn a hand. When I sat down with individual members of the group to explore their personal data, this knowledge did not necessarily automatically transfer. I used a form of smoothing that involved a moving average rather than an average per day. My method was a bit different from the average per day method, but it is something regularly used in reporting financial news so I thought it would not be too exotic. However, doing any smoothing on heart rate data came as a surprise to the community members because smoothing for them was usually a bad thing, a disingenuous thing. Yet there I was, smoothing away so that we could see underlying day-by-day trends of heart rate without having to see the spikes every time they went for a walk. Done on heart rate data instead of pollution data, and done for a different reason, smoothing needed to be introduced again in order to inhabit their imagination of how data parsing worked.

Data Sense was also used in another experimental setting: a group of family caregivers learning to self-track in order to learn something about their own stress. This time in a small group setting, we loaded one volunteer's sleep data into Data Sense and a discussion ensued about his sleep patterns. We discovered using the day-of-week recurring pattern graph that he gets much worse sleep on Wednesday nights, and then he noticed that average duration of sleep seemed low. We toggled back to the sleep data's line graphs and realized they contained a lot of zeros that did not indicate the total absence of sleep but were times he was not wearing the device sensing his sleep. I cleaned out the zeros for him through our feature that enables creating derivative data sets on the fly by selecting points directly on the graph. This participant was not inclined to take the controls and do that cleaning himself, which we had originally designed for. Yet in another sense, this moment was exactly what we designed Data Sense for. He was someone who had no reason to formulate opinions about good data cleaning but saw the need for cleaning because he knew the context of when he was wearing the device. He was able to force the creation of a rendering that was useful in a critical reflection about stress for family caregivers.

For a moment, these tests created their own situatedness. They were a situation that allowed them to contemplate their data, and seeing their data being transformed by someone sitting in the next chair over made a difference. It made a difference to distribute the knowledge between what data needs cleaning and the gestures needed to clean it. In the course of these encounters, my participants got better at seeing what I saw and came to anticipate the next step. As the interviews progressed, the air quality study participant who was initially stumped by the smoothing process had come to request that I toggle between the smoothed and unsmoothed states and then look at it in relation to the air quality data. In these encounters there was no sense in which people were passive. We were having a dialogue in data and making meaning together. It was a tenuous place, a miniature social arena that allowed these stories to emerge. The software was not well positioned to do that on its own, without a social occasion to look and talk. The onboarding mechanisms were necessary, but the real onboarding was taking place in interpersonal connection, a deeper listening. When supported by a social occasion to look and talk, not just a video or pop-up tooltip, a new kind of imaginary was nurtured between a few people for a while.

In these tests, I kept rediscovering over and over again what the Biosensors research program was trying to tell us all along: that it is dangerous business to make assumptions about what people are capable of understanding. In order to get the development underway, we had to stabilize the figure of the end user who explores data. Yet faced with the reality of the heterogeneity of those people and the unpredictability of the stories each one might tell, it grew nearly impossible to establish a definition or baseline about what was and was not intuitive. There were no levels of beginner or advanced, and that made for a tough design process that required us to assert, rather than assume, an assemblage of computational knowledge and an infrastructure for that knowledge that we thought made a contribution to this nascent cultural location. The wide variability in sociocultural understandings, in conceivable data sets and parsings, and in notions of what makes for intuitive data transformations means that no amount of onboarding or in-tool explanation would stabilize meanings appropriately for everyone. They were unstable grounds on which to stand up an infrastructure project.

CONCLUSION

Building an infrastructure for exploratory data analysis reveals gigantic gaps between computational forms and interpretive forms, necessitating new connections between figure and ground that rarely get named explicitly. When two people sit down together to look at data, it does not always look like storytelling. Data set size and groupings need attending to as much as the meaning behind a moving average and why one would use it. We also saw how little of it can be made invisible through building an infrastructure. If derivative data sets are a constant, visualizations need to be transformed, and if data types are unpredictable, choices about these matters cannot be hidden from the data interpreter, even though they do not make for a good story. The underacknowledged, and often undervalued, exploratory work of data was made visible in the act of building Data Sense. The interstitial nature of that work continually exposed yet more areas of software building and testing and yet more areas with gaps between the lifeworlds of computationally dense data sets and our supposedly datafied world, a world that invites us to imagine that data sets could be appropriated but makes it difficult to do in practice.

These yawning gaps in turn raise questions about how data is or is not caught up in social transformation. The real answer to the question of who would use this tool, or one like it, is in the broader social transformations of which the tool is a part. Our goal of elaborating people's capacity to situate data ended in a technical system that could not itself situate data. Perhaps someone who saw society as a thing to be changed by technology would consider that a failure. I do not. It was no coincidence that in the testing, Data Sense did best in group and interpersonal settings like the ones I describe here, where there was interpersonal engagement and socially sanctioned moments to sit down with data, which are remarkably hard to come by even for enthusiastic Quantified Self participants. Indeed, one Quantified Self enthusiast saw immediate value in our tool, by spotting right away that it meant he could now merge the steps data from the various models of fitness monitors he had accumulated over the years and process them as one continuous stream. This is not something that most other users would have spotted at first and indicated he knew something about data work. He noted that finding the time to do that work for his own purposes and not for an employer was extraordinarily difficult, and he knew we had made that work significantly easier for him. No matter how good the tool, the temporal, spatial, and social boundaries that demarcate sanctioned exploration and analysis still matter. Where does time for analysis and reflection fit in the flow of everyday life? Creating a concrete social location for this particular form of thinking is no small matter. Patient support groups, Quantified Self meetups, and community-based participatory research provide glimmers of possibility for such occasions.

One sociotechnical intervention is not a movement. One project cannot change how the cultural logics work and the imaginaries that are ready to hand. A series of interventions, however, is an entirely different matter. The question for this project, and for those who would like data to be used in more socially beneficial ways, is: What alliances are needed for meaningful social transformation to happen? That requires a shared notion of what beneficence looks like, and clearly there is plenty of room for contestation. What at first looks like a beneficial act of valuing the unglorious parts of data work and an act of respect for uncredentialed knowledgeable people could morph over time into increased burdens on more people to do data work (Kragh-Furbo et al. 2016). Indeed, one active area of debate within the air quality group I worked with centered on whether fine-grained data was necessary at all when the evidence was so patently clear that air pollution existed in the area and was contributing to poor health and climate change. A lack of knowledge was not the issue, and there was a risk of getting into a scientific "arms race" of sorts in which underfunded community groups were structurally disadvantaged against industry-funded scientists who would make opposing claims with the same data (see Ottinger 2010). The counterview suggested that community groups already were in that arms race whether they wanted to be or not, and it was better to provoke a change in how the data was collected and parsed rather than allowing others to entirely monopolize those discussions. An analogous discussion was taking place in corners of health self-tracking, too.

If we were to take the view that it is better to gain a foothold in conversations that one cannot stop, then our figure and ground design problem comes full circle. The next ground to open up if data is to benefit movements is a social design problem. Intentional thought needs to be put into shaping the social occasions that knit together the data worlds and lifeworlds of those movements. STEM education is not enough to do that knitting, and neither are the research occasions I share here. As Torben Jensen, Andreas Birkbak, Anders Madsen, and Anders Munk show in chapter 5, it is possible to intentionally design a social occasion in a participatory manner without resorting to social engineering. Part of the puzzle of how data figures in social movements lies in figuring out who does the data work in social movements and how and when that work gets done. In a sense, by doing some of the infrastructure-building work, the need for institutional, social, and cultural infrastructure is revealed in a new way. Such is the unpredictable nature of how making & doing projects travel.

Traveling also can reveal one's own limitations. These social and institutional capacities are not capacities that my research and development lab is well positioned to build. As I write this, we are in discussions with nonprofit organizations that could take on Data Sense. These partners have a specific reason to facilitate the appropriation of data and are better positioned to build the social circumstances in which to do it than we are. In this way, it is not far-fetched to see making & doing as a way of participating in larger social movements. Even if the grounds keep opening up, yet again, as they always seem to, living with the vertigo these openings create also makes it conceivable to at least partially shape the world one would prefer to be living in.

ACKNOWLEDGMENTS

I would like to thank Sangita Sharma, Pete Denman, Lenitra Durham, Lama Nachman, and the rest of the Data Sense team. Gwen Ottinger's collaboration, insight, and friendship helped me to understand what the technology was missing. I also thank Anders Munk, Michael Guggenheim, Gary Downey, and Teun Zuiderent-Jerak for helpful comments and encouragement throughout the writing process.

NOTES

1. “Full stack” refers to someone who can work across both the “back end” of software systems, which is software components that ordinary users do not see, and the “front end,” which is the part of the system that users do see.

2. The “stack” refers to all the hardware and software components necessary to deliver a product or service. The “back end” is considered low in the stack, the front end is high.

3. I am indebted to Suzanne Thomas (pers. comm., November 2017) for pointing out the centrality of work in data systems.

4. The development side of an R&D organization is, to my mind, crucial to enabling making & doing research. Professional developers and computer science students and faculty are not substitutes for one another, even if there are engineering research problems to tackle. Figuring out what kind of coder is necessary for what project is a skill that social scientists would need to develop to do more software-based making & doing research.

5. A minimum viable product is a new product that is sufficiently developed to go into use but not so fully polished that too much investment has gone into it before feedback has been gained from the market.

6. A more theoretical chapter might theorize cultural location in a way that I have chosen not to in order to keep the focus on the specificities of building the software. My point is that, although we could theorize all sociotechnical phenomena as always already emergent, some emergent phenomena are not yet situated comfortably. They neither sit easily in places nor enjoy much recognition. Not all experiences of emergence are quite the same.

REFERENCES

Bowker, Gregory. 2005. *Memory Practices in the Sciences*. Cambridge, MA: MIT Press.

Brown, Phil. 2016. “Learning from Individual Experiences: Measuring Personal Environmental Exposures.” Paper presented at the Personal Environmental Exposures Workshop, National Academy of Science, Washington, DC, November.

Day, Sophie, Celia Lury, and Nina Wakeford. 2014. “Number Ecologies: Numbers and Numbering Practices.” *Distinktion: Scandinavian Journal of Social Theory* 15(2): 123–154.

Dow-Schüll, Natasha. 2016. “Data for Life: Wearable Technology and the Design of Self-Care.” *BioSocieties* 11(3): 317–333.

Dumit, Joseph. 2018. “The Other Ninety Per Cent: Thinking with Data Science, Creating Data Studies—An Interview with Joseph Dumit.” In *Ethnography for a Data-Saturated World*, edited by Hannah Knox and Dawn Nafus, 252–274. Manchester, UK: Manchester University Press.

Gerlitz, Caroline, and Celia Lury. 2014. “Social Media and Self-Evaluating Assemblages: On Numbers, Orderings and Values.” *Distinktion: Scandinavian Journal of Social Theory* 15(2): 174–188.

Guggenheim, Michael, Bernd Kräftner, and Judith Kröll. 2018. “Incubations: Inventing Preventive Assemblages.” In *Inventing the Social*, edited by Noortje Marres, Michael Guggenheim, and Alex Wilkie, 65–93. Manchester, UK: Mattering Press.

Kennedy, Helen, and Giles Moss. 2015. “Known or Knowing Publics? Social Media Data Mining and the Question of Public Agency.” *Big Data and Society* 2(2): 1–14.

Kragh-Furbo, Mette, Adrian Mackenzie, Maggie Mort, and Celia Roberts. 2016. "Do Biosensors Biomedicalize? Sites of Negotiation in DNA-Based Biosensing Data Practices." In *Quantified: Biosensing Technologies in Everyday Life*, edited by Dawn Nafus, 5–26. Cambridge, MA: MIT Press.

Lave, Jean. 1988. *Cognition in Practice: Mind, Mathematics and Culture in Everyday Life*. Cambridge: Cambridge University Press.

Lehtiniemi, Tukka, and Minna Ruckenstein. 2018. "The Social Imaginaries of Data Activism." *Big Data and Society* 6(1): 1–16.

Loukissas, Yanni. 2019. *All Data Are Local*. Cambridge, MA: MIT Press.

Marres, Noortje, Michael Guggenheim, and Alex Wilkie, eds. 2018. *Inventing the Social*. Manchester, UK: Mattering Press.

Nafus, Dawn, ed. 2016. *Quantified: Biosensing Technologies in Everyday Life*. Cambridge, MA: MIT Press.

Nafus, Dawn, and Jamie Sherman. 2014. "This One Does Not Go Up to 11: The Quantified Self Movement as an Alternative Big Data Practice." *International Journal of Communication* 8(11): 1–14.

Ottinger, Gwen. 2010. "Constructing Empowerment through Interpretations of Environmental Surveillance Data." *Surveillance and Society* 8(2): 221–234.

Star, Susan Leigh. 1999. "The Ethnography of Infrastructure." *American Behavioral Scientist* 43(3): 377–391.

Tufte, Edward. 1983. *The Visual Display of Quantitative Information*. Cheshire, CT: Graphics Press.

Van Dijck, José. 2014. "Datafication, Dataism and Dataveillance: Big Data between Scientific Paradigm and Ideology." *Surveillance and Society* 12(2): 197–208.

Verran, Helen. 2001. *Science and an African Logic*. Chicago: University of Chicago Press.

Wheelwright, Jacqueline. 2016. "Self-Tracking for Autoimmune Mastery." https://vimeo.com/144678614, accessed April 10, 2020.

Woolgar, Steve. 1990. "Configuring the User: The Case of Usability Trials." *Sociological Review* 38(S1): 58–99.

The open access edition of this book was made possible by generous funding from Arcadia—a charitable fund of Lisbet Rausing and Peter Baldwin.

The MIT Press would like to thank the anonymous peer reviewers who provided comments on drafts of this book. The generous work of academic experts is essential for establishing the authority and quality of our publications. We acknowledge with gratitude the contributions of these otherwise uncredited readers.

This book was set in Stone by Westchester Publishing Services, Danbury, CT.

Library of Congress Cataloging-in-Publication Data

Names: Downey, Gary Lee, editor. | Zuiderent-Jerak, Teun, editor.
Title: Making & doing : activating STS through knowledge expression and travel / edited by Gary Downey and Teun Zuiderent-Jerak.
Description: Cambridge, Massachusetts : The MIT Press, 2021. | Includes bibliographical references and index.
Identifiers: LCCN 2020044365 | ISBN 9780262539975 (paperback)
Subjects: LCSH: Communication in science. | Communication of technical information. | Science--Study and teaching--Research. | Technology--Study and teaching--Research.
Classification: LCC Q223 .M35 2021 | DDC 507.1--dc23
LC record available at https://lccn.loc.gov/2020044365