Clemson University
TigerPrints

Publications English

12-2016

Big Data, Technical Communication, and the Smart City

Jordan Frith

Follow this and additional works at: https://tigerprints.clemson.edu/english_pubs

Part of the Communication Technology and New Media Commons

Article

Journal of Business and Technical Communication
2017, Vol. 31(2) 168-187
© The Author(s) 2016
Reprints and permission:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/1050651916682285
journals.sagepub.com/home/jbt

Big Data, Technical Communication, and the Smart City

Jordan Frith[1]

Abstract

Big data is one of the most hyped buzzwords in both academia and industry. This article makes an early contribution to research on big data by situating data theoretically as a historical object and arguing that much of the discourse about the supposed transparency and objectivity of big data ignores the crucial roles of interpretation and communication. To set forth that analysis, this article engages with recent discussion of big data and "smart" cities to show the communicative practices operating behind the scenes of large data projects and relate those practices to the profession of technical communication.

Keywords

big data, data analytics, technical communication, smart cities, epistemology, quantification

We live in a world bursting at the seams with data. "Google processes more than 24 petabytes of data per day, a volume that is thousands of times the quantity of all printed material in the U.S. Library of Congress" (Mayer-Schönberger & Cukier, 2013, p. 8); political organizations collect massive

[1]University of North Texas, Denton, TX, USA

Corresponding Author:
Jordan Frith, University of North Texas, 1155 Union Circle #305298, Denton, TX 76203, USA.
E-mail: jordan.frith@unt.edu

amounts of data about voters (Tufekci, 2014b); and sensors embedded in objects produce constant data streams about electricity use, transportation, and infrastructural decay in supposed "smart" cities (IBM, 2014). *Wired* Editor in Chief Chris Anderson (2008) labeled the current moment the "Petabyte Age," but even that descriptor became quickly outdated, and we have already entered the "Zettabyte (2^{70}) Age" (Zikopoulos, Eaton, deRoos, Deutsch, & Lapis, 2012, p. 1). We could try to explain these huge amounts of data to help make sense of our current data deluge, or we could turn to Barnes's (2013) description of the term *big data*: "'Big' as an adjective, then, doesn't get close to describing the size of the data sets now being analysed and manipulated" (p. 298).

Big data has become possibly the biggest buzzword in the sciences and social sciences. According to the research firm Gartner, big data reached the peak of its hype phase in 2013 (Press, 2014), and both industry and academia have embraced big data to increase efficiency (Kitchin, 2014a). But for all the hype, big data is still a phrase that does not have a clear, widely agreed on definition, and many of its proponents have not theoretically situated the big data movement within decades of social theory regarding quantification and epistemology (Barnes, 2013). This article examines big data from a technical communication perspective by first theoretically situating the development of big data and then discussing "smart city" projects to show how the discourses surrounding big data often elide the layers of interpretation and communication that will make big data particularly relevant to the technical communication field.

Providing a critique of big data is not an original contribution to academic literature. Scholars from the disciplines of sociology (boyd & Crawford, 2012; Jurgenson, 2014; Tufekci, 2014b), literary studies (Trumpener, 2009), and geography (Barnes, 2013; Kitchin, 2014a) have already provided excellent critiques of the efficacy of big data approaches and the hidden, positivist assumptions behind the movement. But value can be found in multidisciplinary approaches, and in this article, I examine the claims of big data proponents and critics across disciplinary boundaries. In addition, I do more than just describe big data; I relate the theoretical discussion of big data specifically to technical communication. I show that, through discourses focusing on the objectivity and transparency of large-scale data analysis, the roles of human actors who must interpret and communicate the findings are often rendered invisible.

My main arguments are twofold: First, I argue that the technical communication field must reflect on the epistemological and theoretical lineage and consequences of the big data hype. After all, technical communication researchers are experts in deconstructing and critiquing rhetorics of technology, and few technocratic discourses are more hyped than the push toward big data. And second, I argue that pushing back against the supposed transparency of big data opens up opportunities to identify the crucial role of technical communication inside big data projects. Much of the big data discourse relies on the supposed objectivity of data; however, data must be interpreted and communicated to multiple stakeholders, so practices of technical communication are necessary for the success of big data projects. I am not arguing whether the development of big data is good or bad; it is both, and it is neither. Rather, I am arguing that data never speak for themselves. Someone must always speak for them.

Understanding Data

Data is a word used frequently but rarely defined (Gitelman & Jackson, 2013). According to Rosenberg's (2013) historical textual research, the first use of the word in English came in the 17th century, and the usage exploded with the rationalization and industrialization of the 19th century. The word is a plural of the singular *datum*, and as Gitelman and Jackson argued, "part of what distinguishes data from the more general category, information, is their discreetness. Each datum is individual, separate and separable, while still alike in kind to others in its set" (p. 8). The concept of data has too long a history to fully trace here, but that "discreteness" and what data—big or small—mean and what they can do have changed over time, with the computer revolution of the mid-20th century imparting a new objectivity and calculability to data. Ultimately, regardless of historical time period or epistemological position, "from the beginning, data was a rhetorical concept. Data means—and has meant for a very long time—that which is given prior to argument" (Rosenberg, 2013, p. 36). What I focus on here is how the concept of data as evidence, as a rhetorical basis for decision making, has begun to shift with the move toward big data.

I started with this definition of data before moving on to discuss the *big* for an important reason: to provide historical background on the development of big data and push back against the seeming ahistoricism of the hype. After all, history is filled with claims of information overload and fears about the overwhelming nature of data. Within decades of Gutenberg's invention of the printing press, people were already bemoaning that there were more books than one could possibly read (Hobart & Schiffman, 1998). Data had already become overwhelming, and techniques such as list making exploded in the 16th century as a way to organize textual data in the face of texts' overwhelming growth (Hobart & Schiffman, 1998). Managing data became even more crucial with the development of early capitalism and industrialization. Without time cards and ways to quantify worker output, managing early assembly-line work would have been impossible. And without the quantified urban research of the mid-20th century (Townsend, 2013), the understanding of the city as a "system" would not have happened. To researchers who worked during the explosion of 19th-century industrialization or who worked with some of the first modern censuses, the data must have seemed big indeed.

And then the computing revolution happened. Early computers were built specifically because the amount of data was already too big. These supercomputers—though less powerful than a contemporary iPhone—enabled teams to do calculations either impossible or too time-consuming for humans to do on their own. The early digital computers were ultimately tools for storage and data analysis, tools used to quantify phenomena in new ways. Just as we see discussed in the rhetorics of contemporary big data, these early computers replaced the jobs of humans who had been previously tasked with organizing and analyzing data (Hafner & Lyon, 1998).

Of course, contemporary computers are barely recognizable compared to the early supercomputers of the mid-20th century. And the improvement of computing power and storage, along with the development of the Internet, sensors, radio-frequency identification (RFID) tags, and other data-producing technologies, caused exponentially more data to be produced. People and technologies now produce more analyzable data in a day than they used to produce in a decade. Databases and computing power can now handle data sets that would have been nearly unthinkable even 10 years ago. But even though the writers responsible for much of the hype suggested that big data is a new phenomenon (Anderson, 2008; Mayer-Schönberger & Cukier, 2013), I have provided background that dates back to the 17th century for an important reason: Data have often seemed big and overwhelming, and humans have developed techniques and technologies (from list making to spreadsheeting to digital computing) to make sense of and analyze data in new ways. As is often the case with new "revolutions," much of the hype around big data uses the lack of historical grounding as a rhetorical technique: By imagining big data as something completely new, as its own "big data revolution" (Mayer-Schönberger & Cukier, 2013), the proponents of the movement mask what can be viewed more as a return to traditional positivism than as the creation of a brand new movement (Jurgenson, 2014).

While data collection and analysis have a long history, much of the popular hype around the term big data can be traced to Anderson's

(2008) widely read cover story in *Wired* magazine. The article, provocatively titled "The End of Theory: The Data Deluge Makes the Scientific Method Obsolete," predicted a Kuhnian paradigm shift that would truly mark a departure from earlier forms of quantification. Anderson's main argument was that the scientific method that involves theory, hypotheses, and analysis would soon become outdated with the widespread adoption of big data. He claimed that researchers no longer need models to test through data analysis because it no longer matters why something happens; it only matters that it happens. And by analyzing huge data sets, researchers can identify correlations and then use those findings without understanding why data points correlate or having previously theorized hypotheses. The entire foundation of understanding and manipulating data would be replaced by algorithms and databases.

Although Anderson's article was criticized for being hyperbolic and overstating the case for big data (Barnes, 2013), similar arguments can be found in other writings on the topic. For example, in their book about big data, Mayer-Schönberger and Cukier (2013) suggested that the use of big data signals not just a new method but an epistemological paradigm shift. The authors joined Anderson in arguing that the use of big data represents a new way of understanding the world, one that does not rely on formulating hypotheses, sampling procedures, and establishing causation. Instead, the goal is to collect enough data that the sample does not matter, and with "much more comprehensive data sets we can shed some of the rigid exactitude in a big-data world" (p. 13). Equally important, the authors argued that "society will need to shed some of its obsession for causality in exchange for simple correlations: not knowing why but only what" (p. 7). In other words, researchers could collect huge bodies of data without worrying about traditional sampling procedures, using the big data to identify correlations and then act on the correlations.

Mayer-Schönberger and Cukier (2013) provided many examples of how this method of using big data would work, but a rather mundane yet telling example possibly makes it most clear. In Minneapolis, a man went into a Target and confronted a manager because Target had begun mailing coupons for pregnant women to his 16-year-old daughter (Hill, 2012). The man was angry because he felt that his daughter was too young to receive such coupons, and to the best of his knowledge, she was not pregnant. The man then went home and later found out his daughter was pregnant. So how did Target find out before the father? Because of big data—Target's data analysis showed that women who signed up for baby registries were more likely to buy certain items together. When that man's daughter bought those items

(Target assigns a unique ID to each consumer), Target sent her coupons for baby clothes and cribs. Target had no interest in why pregnant women buy certain items together; all they cared about was using correlations in the data to target advertising.

Anderson (2008), Mayer-Schönberger and Cukier (2013), and other sources positioned the move toward big data as inevitable (Manyika et al., 2011; Smolan & Erwitt, 2012). These works reveal a distinctly deterministic bent, with sources claiming that big data will have impacts similar to those of the Industrial Revolution (Cukier, 2014), that companies will fall behind without big data (Manyika et al., 2011), and that cities must adopt big data to thrive (Dirks, Gurdgiev, & Keeling, 2010). Behind the hype is the implicit belief that, as Anderson claimed, with large enough data sets, "data can speak for themselves," a belief that is by no means unique to big data and was expressed at least as far back as the early positivism of the 19th century (Comte, 1848). As those versed in decades of rhetorical and critical research know, those claims ignore decades of research in the humanities and social sciences (Blyler, 1995; Ceccarelli, 2001; Horkheimer, 1947; Latour, 1987; Law, 2004). In reality, data can never speak for themselves, no matter how big they are (Gitelman, 2013). To claim so is to misunderstand the notion of data. As Bowker (2005) argued, "Raw data is both an oxymoron and a bad idea; to the contrary, data should be cooked with care" (pp. 183–184).

But that cooking—with "recipes" for collecting, structuring, interpreting, and acting on data—is often elided in many of the more popular discourses about big data. In an ideal world, big data analyses would be able to study a phenomenon in its entirety. In reality, data sets can almost never capture everything, and the belief that they can occasionally has pernicious effects. Take, for instance, data produced in the smart city plans (Kitchin, 2014b). Some cities rely on social media data analysis to identify infrastructure that needs to be improved. But as Crawford (2013) pointed out, not everyone uses social media, and relying on big data approaches to urban planning can lead to analyses that leave people out of the discussion. The same applies to the use of social media data to predict political sentiment or disease outbreak (Tufekci, 2014a). And these examples apply to more than social media. Data do not come from nowhere. Not all data can be collected. Researchers and companies choose what data to collect—where to deploy sensors and tags, what to follow about people's online browsing, which social media sites to research, which employee interactions to track—and that collection often comes with biases, and those biases might be different from those that come from traditional research sampling. As Williams

(2013) argued, "a data set is already interpreted by the fact that it is a set: some elements are privileged by inclusion, while others are denied relevance through exclusion" (p. 41).

In effect, while much of the big data hype positions data collection as a neutral process, choosing how, when, and where to collect data is always a limiting factor in any analysis. This limitation points to the human actors behind the design of big data projects. It also relates to another significant push back against the supposed objectivity of big data: No matter how big the data set and how advanced the algorithm, findings still require human actors to interpret them. In a TED Talk on big data, Cukier (2014) posited that one of the major impacts of the big data revolution is that companies will need far fewer people to interpret and act on findings. Instead, analyses will be able to reveal correlations and provide outputs free of human bias. Similar arguments are found in much of the big data hype, with Mayer-Schönberger and Cukier (2013) writing that "we don't always need to know the cause of a phenomenon. We can let data speak for itself" (p. 14). The idea of removing human bias from the collection and analysis of data is not a new one; the bedrock of traditional positivism is that data could be used to analyze the world as it truly is. And as rhetorical scholars have shown (Gross, 1991; Miller, 1979), the rhetoric of science works hard to remove human actors from how scientific results are presented, using writing techniques such as passive voice and third person as a way to "let data speak for itself." But the "view from nowhere" approach ignores the role that human actors play in choosing what to collect and then how to interpret and act on the correlations found in data analyses (Jurgenson, 2014).

This criticism of the position that data collection and analysis can be free from human bias is particularly important because big data approaches have repeatedly turned up spurious correlations (Kobielus, 2014). In contrast to much of the hype, big data still requires researchers to interpret correlations found in the data, and as Kitchin (2014a) argued, that will not happen successfully without subject-matter experts. A huge data set filled with multiple fields will turn up many correlations; with larger data sets, there are more chances to identify correlations. But most correlations do not matter. For example, if the society of technical communication assembled a huge database with 50 fields about technical communication professionals that included both personal and professional information, an analysis might find that dog ownership correlates with higher salaries. But that correlation is meaningless, and just relying on correlation strength may lead researchers to ignore weaker correlations that have far greater explanatory power. As boyd and Crawford (2012) argued, big data approaches lend themselves to this kind of *apophenia*, which refers to "seeing patterns where none actually exist, simply because enormous quantities of data can offer connections that radiate in all directions" (p. 668). Without human actors to interpret data analyses, then such approaches will often turn up these types of spurious correlations.

Criticizing the notion that any amount of data or any statistical analysis could remove human actors from the equation is important for the profession of technical communication. As I cover in more detail through the example of smart cities, discourses that elide human actors ignore the many layers of communication that go into making big data accessible and actionable. Acknowledging this criticism can help us begin recognizing how important various skills of technical communication will be to the success or failure of big data projects.

Big Data and Technical Communication

Taken to its logical extreme, the widespread embrace of big data may have pernicious effects on the profession of technical communication. Technical communicators must often work with subject-matter experts to interpret and explain data (Jeyaraj, 2004). If the data in big data approaches could truly represent the world completely and remove human bias and interpretation from the equation, technical communicators would be less important. Companies would no longer need anyone to turn data into accessible narratives because the findings would be self-explanatory. But those assumptions fundamentally misunderstand what data are and what they can do.

First off, data are not objective entities that can be collected free of bias, nor do data construct their own narratives. That is true in academic research settings, and it is true in corporate settings. Data are instead rhetorically constructed to make meaning (Latour, 1987). They always have been and always will be even as data sets become exponentially larger. But regardless of the criticisms of the movement, big data has become mainstream. Analysts expect there to be over a million jobs in the coming years in big data–related fields (Pettey, 2012), companies have poured significant resources into data collection and analysis (Smolan & Erwitt, 2012), and research-funding agencies have increased funding streams for big data projects (Sawyer, 2008). These jobs will exist, many big data projects will continue to be funded, and our field must recognize and be ready to argue for how technical communication skills are necessary for the success of big data.

I began this article with a lengthy discussion of the theoretical underpinnings of data and critiques of big data hype to show why technical communication skills are necessary. In critiquing the hype, I also suggested how the emphasis on objectivity and data analysis ignored communicative aspects of big data. To make that theoretical discussion more concrete, the next section analyzes a wide-ranging example of the instantiation of big data: the growth of "smart" cities. My goal is to use smart cities as an extended example to show how the rhetorics of big data often elide the many layers of communication necessary for successfully implementing big data.

When Data Cannot Speak for Themselves: Big Data and the Growth of Smart Cities

A prominent example of the deployment of big data has been the growth of smart cities. The term *smart cities* is broad and resists simple definition, but it mostly refers to the use of digital technologies to produce data to increase cities' efficiency, improve their livability, and promote their safety. The most visible company behind the push toward smart urban technology has been IBM, which in the early to mid-2000s moved away from PC hardware to focus on urban technologies through its trademarked Smarter Cities campaign (Söderström, Paasche, & Klauser, 2014). For my purposes, I use the phrase *smart city* to refer to data-driven urban projects in general and *smarter cities* to specifically refer to IBM's marketing materials.

Smart cities are a microcosm of the larger discussion of big data traced throughout this article. They are positioned as the future, with IBM (2014), for example, arguing that cities must face the challenges of the 21st century by embracing smart technologies and data analysis. So far the industry has been highly successful commercially, with IBM already creating billions of dollars of new smart infrastructure and Cisco pouring money into its "Internet of Everything for Cities" projects (Kitchin, 2014b). In effect, smart city campaigns are "conceived to channel urban development strategies through the technological solutions of IT companies" (Söderström et al., 2014, p. 308).

The development of smart urban infrastructure cannot be divorced from the growth of big data. As the IBM Smarter Cities campaign (2014) states, urban "leaders see transformative possibilities in using big data and analytics for deeper insights." IBM's point about cities and data is seconded by prominent urban leaders, with former New York City Mayor Michael Bloomberg (2014) stating that the urban "revolution is our growing ability to use data to improve the services that government provides" and that "if you can't measure it, you can't manage it. And I brought that approach with me from the private sector to New York's city hall. Our administration looked for ways to use data—and to collect more data—to help us better serve New Yorkers" (p. v). The future of cities, ranging from the smart electrical grid that determines power use, to data-driven policing, to decisions about infrastructural repair, will supposedly be driven by big data analytics.

The vast majority of the smart city hype focuses on technologies (Greenfield, 2013; Kitchin, 2014b). Cities will build huge data centers, use improved analytics, and collect data from sensor and RFID technology, social media, and legacy data sources. What almost none of the smart city promotional literature addresses is just how people will interpret and act on these data. The human actors behind urban big data projects are often rendered invisible in campaigns such as IBM's Smarter Cities or more technocratic discussions of smart technology (Smart Cities USA, 2015). In effect, smart city marketing materials are closely tied to the big data hype that purports that data can speak for themselves. The actors, just as Cukier (2014) warned about the "big data revolution," are rendered absent by the technocratic focus on smart technology and data analysis rather than human interpretation and communication.

And for that reason, the development of smart cities—as just one of many instantiations of big data—exemplifies why the epistemology of big data is particularly relevant to the field of technical communication. These human actors who are rendered invisible by the hype about data being free from interpretation or by smart technologies and databases are often people engaged in fundamental practices of technical communication. The popular conception of the smart city as a primarily technological, data-driven fix to urban problems ignores the many layers of interpretation, communication, and visualization necessary for any big data project to succeed. In these discourses about the transparency of data, the communicative practices of technical communication are sacrificed at the altar of positivist quantification. To see the levels of communication that are implied but often explicitly ignored in smart city materials and, by extension, big data hype, we can look to many examples of smart cities that show why practices of technical communication (e.g., information architecture, data and actionable narrative, usability testing, help documentation, and visualization) are all still necessary even with the growth of smart technologies.

To start, one of the major problems threatening to slow the growth of big data is the difficulty in structuring large data sets. Data projects are often expensive, and a significant percentage of such projects fail (Gane, Venn, & Hand, 2007). They do so in part because it is difficult to organize databases in ways that enable accurate data analysis. Smart city projects face the same problems, and as Michael P. Flowers, former chief analytics officer of New York City, explained, one of the main challenges he faced was dealing with legacy data sources and connecting previously discrete city databases (Goldsmith & Crawford, 2014, Chapter 6). The structural issues that make legacy data difficult to deal with or prevent databases from interconnecting are aided by a well-developed metadata system, an area with which many technical communicators are already familiar (Andersen, 2014; Giordano, 2013; Goolsby, 2012; McCarthy, Grabill, Hart-Davidson, & McLeod, 2011; Panke & Gaiser, 2009). In fact, technical communication researchers have argued that the field is increasingly moving toward forms of information architecture (Salvo, 2004), and big data approaches like the ones championed by New York City's Office of Data Analytics have shown how necessary metadata and information architecture are to smart urban projects. Huge bodies of data are fairly worthless unless they are formatted in a way that can be analyzed by algorithms. No matter how much smart city projects may focus on sensor technology, social media data collection, or smart grid output, data do not format or organize themselves. Such tasks require professionals with sophisticated understandings of both metadata and database structure to succeed.

Of course, formatting existing and new data is only an early step in any big data analysis. The findings must also be explained to various stakeholders: "While some data analysis is ceded to algorithms, especially in the grunt work of processing and calculating, direction and interpretation is still largely the preserve of a human analyst" (Kitchin, 2014a, p. 160). This factor is mostly elided in prominent smart city campaigns such as IBM's Smarter Cities. Discussions of smart cities rarely talk about how public employees will actually use the data collected to make decisions, and there are few critical reflections on where the data come from and what they may be missing. In effect, the focus on smart data ignores that people will still have to turn these data into something meaningful to inform city planning decisions, a point that is central to the field of technical communication. After all, as Pflugfelder (2013) argued, one of technical communicators' greatest strengths is their ability to "produce coherent and meaningful narratives from data" (p. 19).

An example from Townsend's (2013) book on smart cities clearly shows the role that technical communicators (or at least people with technical communication skills) will be required to play in interpreting and explaining big data findings. In 2011, IBM worked with the city of Portland, Oregon, to develop a "system dynamics for smarter cities." This system "wove together more than three thousand equations" (p. 82) based on data produced about Portland, allowing city planners to search for correlations and model various potential policy changes. Townsend provided details of the benefits and drawbacks of the system, and most important, he warned that "a far bigger risk is that public officials will accept the advice of these black boxes unquestioningly" (p. 88). He quoted one project member's fear that Portland's mayor would blindly trust the data to "tell him what the right thing to do was." Various members of the project then had to clearly explain the limitations of big data to the mayor to "make sure that he understood that models aren't oracles" (p. 89). They also had to explain the meaning of different findings to policy makers so that they might understand which correlations are meaningful and which are not. In Townsend's description, the big data project in Portland directly contrasts with the hype about letting data speak for themselves. Instead, a crucial piece of the project involved data-savvy communicators' explaining findings and limitations to multiple stakeholders.

Smart cities also feature many nods to "open data" that are shared with the public (Deakin, 2012). But here again we see that cursory discussions of "open data" and "open government" often ignore that data in their raw form are often meaningless. The goal of open data is to make the public more informed about decision making, but people who have enacted open government projects have found that simply providing data is not enough (Goldsmith & Crawford, 2014). The data must be explained and formatted so that they are accessible. For example, Daniel O'Neil, executive director of the Smart Chicago Collaborative, worked with the city of Chicago on the site schoolcuts.org, which was intended to use open data to "help parents understand what schools were being closed and why" (Goldsmith & Crawford, 2014, p. 90). O'Neil knew there could be problems with how the data were being presented, so he decided to usability test the system. He invited resident parents in to use the interface and found they could not make sense of the data formatting. His organization then took the results of the usability test into account and changed the types of data presented to residents and how the data were visualized. Simply crunching huge data sets did little to help residents better understand why some schools were being closed. Instead, through testing and considering the target audience, the city was able to create new data visualizations that could be understood and acted on by the audience.

Another example of the links between practices of technical communication and urban open data is the City of Austin Texas's open-data portal (https://data.austintexas.gov/). The open-data portal works as the civic hub for a wide range of data sets about the city. The project was created at the ATX Hack for Change event in 2014, and the Web site is maintained by volunteers of the Open Austin Project. The actual data sets are formatted and organized by the open-data company Socrata, which labels itself as the "global leader in software solutions that are designed exclusively for digital government" (O'Neil, 2015). One of Socrata's major areas of expertise focuses on open-data portals, and the company has helped develop portals for cities besides Austin, including Dallas, Seattle, and New Orleans. Each city's portal features a different interface and different data sets, but they all share a focus on usability and fundamental practices of technical communication. A deeper look at the Austin portal in particular clearly shows this focus.

Two of the first things we see when we go to the Austin open-data portal are a scrolling banner featuring video tutorials about how to use the portal and a link to a How-To Wiki (see Figure 1). The instructional wiki is a guide to the portal and features an extensive set of user documentation that instructs citizens on how to navigate data sets, create a map, create an account, and filter, visualize, and contribute data. In sum, the wiki features more than 10 pages of help documentation about the open-data portal and is a key part of the project's success. Few examples could more clearly show how fundamental practices of technical communication, in this case, textual and video help documentation, are necessary for the success of big data and open-data projects.

Like the Chicago open-data project, the Austin data portal works more efficiently because it is designed with something that often remains unaddressed in smart city literature: audience. The open-data portal is designed explicitly for a specific type of citizen who is tech savvy enough to manipulate a fairly intuitive interface and consult documentation but does not necessarily have statistical expertise. In fact, on the first page of the How-To Wiki, the writers directly address intended audience in a warning to statistical experts that the open-data portal does not likely offer the tools they need:

> If you think in spreadsheet or are fluent in statistical software, you may be frustrated with the analytical tools Socrata provides. It may be more fruitful for you to just download a data set and analyze it in your tool of choice than learn how to use the tools baked into Socrata. The tools here have limitations, and you'll be able to accomplish more in a tool you know well.

Rather than tech experts, the target audience consists of everyday citizens, and the portal provides data sets that cover a wide range of topics, from

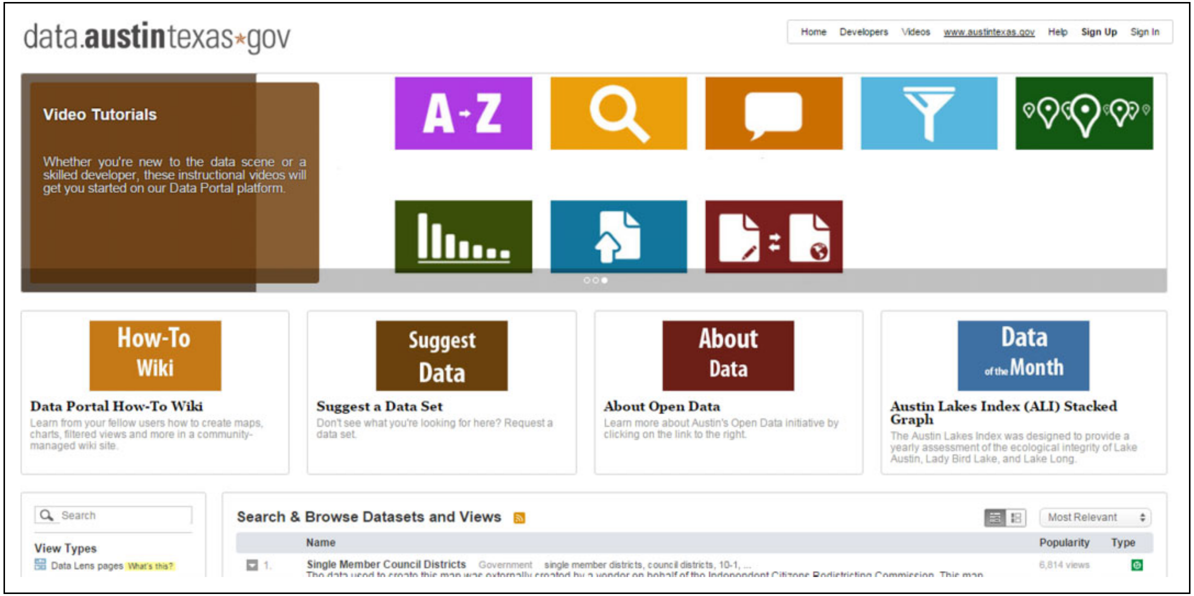

Figure 1. The home page of Austin's open-data portal.

maps of city golf courses to restaurant inspection scores to a map of declared dangerous dogs. If users click on one of these data sets, they are offered different options depending on the topic, including maps of the spatial distribution of data points, filters that enable conditional formatting, and visualizations that range from using different map features to displaying the data in multiple graphing formats. By no means are all of the data particularly easy to understand, and certain sets have limited options, but the How-To Wiki combined with the video tutorials and clean user interface provide citizens with the opportunity to use and contribute to the data portal. Thus, it is not the data sets in themselves that help citizens better understand their city; it is the data combined with the documentation, visualizations, and filters that make the portal valuable to citizens trying to navigate Austin's civic landscape.

Just in this brief discussion of smart cities, we can see the many layers of technical communication involved in big data projects. In New York City, the chief analytics officer's legacy-data project required multiple levels of information architecture expertise. In Portland, the IBM project required communication experts to turn data analysis into actionable narratives. In Chicago, the open-data schoolcuts.org required usability testing the open-data site and creating new visualization when the audience could not make sense of the first iteration of the data display. And in Austin, the open-data portal required extensive help documentation and usable data visualizations to help residents better understand their city. All of these examples push back against the rhetoric of big data speaking for themselves. In smart cities alone, we can see the many layers of communication expertise necessary to get big data projects off the ground, and these examples are important reminders that big data will involve not only data scientists: They will also involve essential skills of technical communication as a crucial piece of the larger puzzle.

Conclusion

Big data projects are growing and will continue to grow. Even if they never live up to their hype, such projects will have an impact on the workplace, and analysts expect a significant increase in data-related jobs in the coming years. Manyika et al. (2011) estimated that the "United States alone faces a shortage of 140,000 to 190,000 people with deep analytical skills as well as 1.5 million managers and analysts to analyze big data and make decisions based on their findings" (p. 3). In some cases, as Cukier (2014) warned, big data might replace some professionals, just as many technologies have automated processes in the past. But as this discussion of smart cities has shown, big data projects will often not replace the need for people with technical communication skills.

To make that case, I began this article with a theoretical discussion of data and a critical analysis of some of the popular discourses surrounding the "big data revolution." I argued that our field should push back against the belief that data can speak for themselves in an objective manner free from bias. A study by Shah, Horne, and Capella (2012) revealed why understanding the implications of these discourses is so important. Their study examined how 5,000 employees at global firms understood data analytics. The researchers found that 43% were "unquestioning empiricists" who trusted whatever the data told them, 19% were "visceral decision makers" who went with their instincts over data, and only 38% were "informed skeptics" who recognized the value of data analytics while also recognizing their limitations. Those findings are troubling because those who are best suited to work with data without relying on them completely are outnumbered by those who let the data speak for themselves and those who ignore data altogether. Workplaces, whether inside smart cities or other big data projects, will need people who can communicate effectively to interpret which findings are meaningful, transform data analysis into meaningful narratives, and work with stakeholders to act on data.

In conclusion, my two main arguments have been (a) that our field must reflect on the epistemological and theoretical consequences of the big data hype and (b) that recognizing the ways in which big data discourses render invisible the necessary levels of communication helps us identify how technical communication skills fit within these projects. By combining theories of data with the more concrete examples of smart city projects, I have shown why our field and our profession have much to contribute to the growth of big data. After all, no matter how big our data get, they can never speak for themselves. They will always need someone to speak for them.

Declaration of Conflicting Interests

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

Funding

The author received no financial support for the research, authorship, and/or publication of this article.

References

Andersen, R. (2014). Rhetorical work in the age of content management: Implications for the field of technical communication. *Journal of Business and Technical Communication*, *28*, 115–157.

Anderson, C. (2008). The end of theory: The data deluge makes the scientific method obsolete. *WIRED*. Retrieved November 6, 2014, from http://archive.wired.com/science/discoveries/magazine/16-07/pb_theory

Barnes, T. J. (2013). Big data, little history. *Dialogues in Human Geography*, *3*, 297–302.

Bloomberg, M. (2014). Foreword. In Stephen Goldsmith & Susan Crawford (Eds.), *The responsive city: Engaging communities through data-smart governance* (pp. iii–vi). New York, NY: Jossey-Bass.

Blyler, N. R. (1995). Research as ideology in professional communication. *Technical Communication Quarterly*, *4*, 285–313.

Bowker, G. (2005). *Memory practices in the sciences*. Cambridge, MA: MIT Press.

boyd, d., & Crawford, K. (2012). Critical questions for Big Data. *Information, Communication & Society*, *15*, 662–679.

Ceccarelli, L. (2001). *Shaping science with rhetoric: The cases of Dobzhansky, Schrodinger, and Wilson*. Chicago, IL: University of Chicago Press.

Comte, A. (1848). *A general view of positivism*. Cambridge, England: Cambridge University Press.

Crawford, K. (2013). The hidden biases in big data. *Harvard Business Review*. Retrieved June 12, 2014, from http://blogs.hbr.org/2013/04/the-hidden-biases-in-big-data/

Cukier, K. (2014). Big data is better data. *TED*. Retrieved from http://www.ted.com/talks/kenneth_cukier_big_data_is_better_data?language=en

Deakin, M. (2012). Intelligent cities as smart providers: CoPs as organizations for developing integrated models of eGovernment Services. *Innovation: The European Journal of Social Science Research*, *25*, 115–135.

Dirks, S., Gurdgiev, C., & Keeling, M. (2010). Smarter cities for smarter growth. *IBM*. Retrieved from http://www-01.ibm.com/common/ssi/cgi-bin/ssialias?infotype=PM&subtype=XB&appname=GBSE_GB_TI_USEN&htmlfid=GBE03348USEN&attachment=GBE03348USEN.PDF#loaded

Gane, N., Venn, C., & Hand, M. (2007). Ubiquitous surveillance: Interview with Katherine Hayles. *Theory, Culture & Society*, *24*, 349–358.

Giordano, C. (2013). Navigating the road to a business-driven data conversion strategy. *TechWhirl*. Retrieved January 4, 2015, from http://techwhirl.com/navigating-road-to-data-conversion-strategy/

Gitelman, L. (Ed.). (2013). *Raw data is an oxymoron*. Cambridge, MA: MIT Press.

Gitelman, L., & Jackson, V. (2013). Introduction. In L. Gitelman (Ed.), *Raw data is an oxymoron* (pp. 1–14). Cambridge, MA: MIT Press.

Goldsmith, S., & Crawford, S. (2014). *The responsive city: Engaging communities through data-smart governance*. New York City, NY: Jossey-Bass.

Goolsby, C. (2012). Tech comm, data conversion and the entrepreneur's evolution. *TechWhirl*. Retrieved January 4, 2015, from http://techwhirl.com/technical-communication-data-conversion-and-the-entrepreneurs-evolution/

Greenfield, A. (2013). *Against the smart city*. New York City, NY: Do Projects.

Gross. (1991). Does rhetoric of science matter? The case of the floppy-eared rabbits. *College English*, *53*, 933–943.

Hafner, K., & Lyon, M. (1998). *Where wizards stay up late: The origins of the Internet*. New York, NY: Simon and Schuster.

Hill, K. (2012). How Target figured out a teen girl was pregnant before her father did. *Forbes*. Retrieved November 6, 2014, from http://www.forbes.com/sites/kashmirhill/2012/02/16/how-target-figured-out-a-teen-girl-was-pregnant-before-her-father-did/

Hobart, M., & Schiffman, Z. (1998). *Information ages: Literacy, numeracy, and the computer revolution*. Baltimore, MD: John Hopkins University Press.

Horkheimer, M. (1947). *Eclipse of reason*. New York, NY: Martino Fine Books.

IBM. (2014). Smarter cities. *IBM*. Retrieved from http://www.ibm.com/smarterplanet/us/en/smarter_cities/overview/

Jeyaraj, J. (2004). Liminality and othering: The issue of rhetorical authority in technical discourse. *Journal of Business and Technical Communication*, *18*, 9–38.

Jurgenson, N. (2014). View from nowhere. *The New Inquiry*. Retrieved November 7, 2014, from http://thenewinquiry.com/essays/view-from-nowhere/

Kitchin, R. (2014a). *The data revolution: Big data, open data, data infrastructures and their consequences*. London, England: Sage.

Kitchin, R. (2014b). The real-time city? Big data and smart urbanism. *GeoJournal*, *79*, 1–14.

Kobielus, J. (2014). Big data's bogus correlations. *IBM*. Retrieved November 7, 2014, from http://www.ibmbigdatahub.com/blog/big-datas-bogus-correlations

Latour, B. (1987). *Science in action: How to follow scientists and engineers through society*. Cambridge, MA: Harvard University Press.

Law, J. (2004). *After method: Mess in social science research*. New York, NY: Routledge.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., … Byers, A. (2011). *Big data: The next frontier for innovation, competition, and productivity*. Mckinsey Global Institute. Retrieved from http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation

Mayer-Schönberger, V., & Cukier, K. (2013). *Big data: A revolution that will transform how we live, work, and think*. New York, NY: Eamon Dolan/Houghton Mifflin Harcourt.

McCarthy, J. E., Grabill, J. T., Hart-Davidson, W., & McLeod, M. (2011). Content management in the workplace: Community, context, and a new way to organize writing. *Journal of Business and Technical Communication*, *25*, 367–395.

Miller, C. (1979). A humanistic rationale for technical writing. *College English*, *40*, 610–617.

O'Neil. (2015). Socrata launches data-as-a-utility™ cloud platform for 21st century digital government. *Socrata*. Retrieved from http://www.socrata.com/newsroom-article/socrata-launches-data-as-a-utility-cloud-platform-for-21st-century-digital-government/

Panke, S., & Gaiser, B. (2009). "With my head up in the clouds": Using social tagging to organize knowledge. *Journal of Business and Technical Communication*, *23*, 318–349.

Pettey, C. (2012). Gartner says big data creates big jobs: *Gartner*. Retrieved November 7, 2014, from http://www.gartner.com/newsroom/id/2207915

Pflugfelder, E. H. (2013). Big data, big questions. *Communication Design Quarterly Review*, *1*, 18–21.

Press, G. (2014). It's official: The Internet of things takes over big data as the most hyped technology. *Forbes*. Retrieved September 19, 2014, from http://www.forbes.com/sites/gilpress/2014/08/18/its-official-the-internet-of-things-takes-over-big-data-as-the-most-hyped-technology/

Rosenberg, D. (2013). Data before the fact. In L. Gitelman (Ed.), *Raw data is an oxymoron* (pp. 15–40). Cambridge, MA: MIT Press.

Salvo, M. J. (2004). Rhetorical action in professional space: Information architecture as critical practice. *Journal of Business and Technical Communication*, *18*, 39–66.

Sawyer, S. (2008). Data wealth, data poverty, science and cyberinfrastructure. *Prometheus*, *26*, 355–371.

Shah, S., Horne, A., & Capella, J. (2012). Good data won't guarantee good decisions. *Harvard Business Review*, *90*, 23–25.

Smart Cities USA. (2015). *Smart America*. Retrieved from http://smartamerica.org/teams/smart-cities-usa/

Smolan, R., & Erwitt, J. (2012). *The human face of big data*. New York, NY: Sterling.

Söderström, O., Paasche, T., & Klauser, F. (2014). Smart cities as corporate storytelling. *City*, *18*, 307–320.

Townsend, A. (2013). *Smart cities: Big data, civic hackers, and the quest for a new utopia*. New York, NY: Norton.

Trumpener, K. (2009). Paratext and genre system: A response to Franco Moretti. *Critical Inquiry*, *36*, 159–171.

Tufekci, Z. (2014a). Big questions for social media big data: Representativeness, validity and other methodological pitfalls. In *Proceedings of the 8th International AAAI Conference on Weblogs and Social Media* (n.p.). Ann Arbor, MI: ICSWM.

Tufekci, Z. (2014b). Engineering the public: Big data, surveillance and computational politics. *First Monday*, *19*. Retrieved from http://firstmonday.org/ojs/index.php/fm/article/view/4901/4097

Williams, T. D. (2013). Procrustean Marxism and subjective rigor: Modern arithmetic and its readers. In L. Gitelman (Ed.), *Raw data is an oxymoron* (pp. 41–60). Cambridge, MA: MIT Press.

Zikopoulos, P. C., Eaton, C., deRoos, D., Deutsch, T., & Lapis, G. (2012). *Understanding big data*. New York, NY: McGraw Hill.

Author Biography

Jordan Frith is an assistant professor of technical communication at the University of North Texas. His research focuses on the social impacts of mobile and social media, and he has published multiple books and journal articles examining the intersections of emerging media and people's experience of space and place. His newest book—*Smartphones as Locative Media*—was published in 2015.