A Round-Table Discussion of “Big” Data in Qualitative Organizational Communication Research

Ryan S. Bisel¹, J. Kevin Barge², Debbie S. Dougherty³, Kristen Lucas⁴, and Sarah J. Tracy⁵

Forum Introduction by the Editor

The forum guest editor Ryan Bisel in this issue takes on the topic of big data and presents a round table that grew out of a conference panel. Five scholars engage in a discussion of the social and cultural trend of big data and implications to qualitative organizational communication research. The contributors respond to questions and delve into a number of issues, from theoretical, to institutional, to operational, to practical, by sharing thoughts and experiences about definition, assumptions, theory building, execution at every stage of a big data project and reflections beforehand and afterward.

Opening Remarks

The phrase “big data” refers to a trend in corporate and academic circles to utilize increasingly available stores of structured and unstructured

¹University of Oklahoma, Norman, USA
²Texas A&M University, College Station, USA
³University of Missouri, Columbia, USA
⁴University of Louisville, KY, USA
⁵Arizona State University, Tempe, USA

Corresponding Author:
Ryan S. Bisel, Department of Communication, University of Oklahoma, 610 Elm Ave., Rm. 101, Norman, OK 73019, USA.
Email: ryanbisel@ou.edu
information for improving decision making and bolstering new knowledge creation. Yet the term “big data” is also “a meme and a marketing term” (Lohr, 2012, para. 7) in the sense that the phrase represents underlying cultural assumptions that more data should improve our lives, organizations, and society (McAfee & Brynjolfsson, 2012). As a cultural meme and movement (Sardana, 2013), big data have the potential to shape widely held expectations about how research ought to be conducted and what kinds of research are worthwhile. The big data trend motivated some social scientists, writing in the journal Science, to declare the emergence of a new field (i.e., computational social science, Lazer et al., 2009) and U.S. universities are increasingly seeking and receiving government grants and awards for big data research projects (DeSantis, 2012). However, big data approaches are not without their detractors and pitfalls (Parry, 2014; Pullum, 2013). Puschmann and Burgess (2014) explained that scientific rhetoric can invoke big data to bolster authority from “the sheer abundance of information available”; meanwhile, critics argue that “big data poses significant methodological challenges, at times trading large scale for reduced depth” (p. 1691).

What does the big data trend mean for qualitative organizational communication research? At the 2013 National Communication Association (NCA) annual meeting in Washington, D.C., five experienced organizational communication scholars met to discuss the role of big data in qualitative research. The panelists included (alphabetically) J. Kevin Barge (Texas A&M University), Ryan S. Bisel (University of Oklahoma), Debbie S. Dougherty (University of Missouri), Kristen Lucas (University of Louisville), and Sarah J. Tracy (Arizona State University). Topics ranged from paradigmatic concerns regarding the underlying assumptions of the big data movement to the pragmatics of dealing with large unstructured, qualitative datasets. What follows is a continuation of those panel presentations and discussions. The panelists offer insights as methodologists interested in theories of knowing as well as insights from their professional research experiences with managing large qualitative projects. In this way, the big data metaphor provided a conversational trigger and anchor, but also allowed the researchers to define and redefine “big” as it relates to qualitative research in organizational communication. The conversation is organized around a series of questions regarding assumptions embedded in ideas about big data, grant-funded research, research design, data collection, and data analysis in qualitative organizational communication research.

Bisel: I want to start our conversation by asking for your thoughts on the role of big data in qualitative organizational communication research. How important is the amount of qualitative data collected?
Dougherty: I would like to begin by defining big qualitative data. Of course, there are some similarities with big quantitative data, but the differences are important. Big quantitative data are usually preexisting or “found” data; in other words, researchers take advantage of large quantities of data collected by others, often corporations or governmental organizations. These data help quantitative scholars identify small but meaningful differences that would not be apparent in smaller studies. It also provides researchers with access to populations they may otherwise not be able to access. Unfortunately, because the data are found, researchers make do with whatever design was implemented, making most big quantitative data limited for academic development. Nonetheless, the idea of big data garnered a lot of buzz and enthusiasm. For me, what constitutes big qualitative data is more complicated. I think of big qualitative data as not only characterized by size but also complexity. Big qualitative data can be defined as (a) a lot of data, (b) highly complex data involving multiple points of triangulation, or (c) a complex data analysis process that provides unique and unusually deep insight. Big qualitative data can be found or created; however, if a big qualitative dataset is found, it is plagued by problems similar to those experienced by quantitative scholars who use big data.

For me, the real issue is usefulness: Either big data are useful or they are not. A big dataset that is just big, which does not add anything new past the point of conceptual saturation is of no greater value than a small and carefully designed study—such studies are of no greater value to either the researcher or the reader. Using big data simply because it is easy and the researcher does not have to do interviews or transcriptions will probably result in findings that are of questionable value. Instead, researchers need to ask themselves what they can do to answer provocative research questions. Let me offer two examples: A number of years ago, one of my graduate advisees, Francie Smith, decided she wanted to study retirement for her dissertation work. After a long period of immersion in the literature, she decided to study the organizational socialization of retirement. She was particularly intrigued with combining Jablin’s model of socialization with retirement discourses. The conversation went something like this:

**Debbie:** Francie, if you want to explore the socialization of retirement then you will need to talk to people in each of the socialization phases.

**Francie:** Oh. Okay.

**Debbie:** You will need to saturate each of those categories.

**Francie:** Oh. Wow. I can see that.
Debbie: And you will need to make sure that you consider people from a range of social classes and other differences.

Francie: I have considered that. I also want to have the participants bring in an object that symbolizes retirement.

Debbie: Why is that?

Francie: I read a study where the author did something similar and thought it created a really interesting conversation.

Debbie: How will you analyze it?

Francie: I have not decided yet.

Note the increasing complexity in the design process. The nature of the question made it necessary for Francie to conduct multiple sets of interviews, each set of interviews accounting for the diversity of the workforce. To achieve saturation, Francie conducted more than 80 interviews. She then did a comparative analysis, both within and between each group. This was an amazing dissertation and resulted in publication that has been well received (Smith & Dougherty, 2012). This is a big data study, not because Francie wanted to do 80 plus interviews, but because it was necessary for her to conduct this many interviews to achieve her goal. Similarly, Jenny Dixon conducted more than 60 in-depth interviews because she wanted to explore the experience of sexuality in the workplace. There are many types of sexualities, necessitating that she interview a wide range of participants. This was a big data study too, not because Jenny had any desire to conduct a never-ending array of interviews, but because her dissertation questions required that she do so. Again, this work resulted in a high-quality publication (Dixon & Dougherty, 2014).

Lucas: I agree with Debbie wholeheartedly that big qualitative data are about size and complexity. But there is a troubling tendency to think purely about quantities of data instead of other qualitative markers of bigness. We ask questions such as the following: How many interviews were conducted? Pages transcribed? Hours observed? Field notes taken? Documents gathered? Hannah and Lautsch (2011) call this approach to bigness credentialing counting, or the counting that is done for the purpose of bolstering confidence in the findings of a qualitative study. Given how much attention these counts are paid throughout the review process, how much space is dedicated to them within our articles, and the trend for the numbers reported to keep growing in published studies, it is easy to believe that the size of a data collection is the “bigness” that matters most. But if we want to have the biggest impact as qualitative organizational communication researchers, we should
care less about the numbers and more about big thinking. It is not that
the size of qualitative datasets is unimportant. Indeed, significant data
collections are a cornerstone of good qualitative research. But a bigger
quantity of data—particularly for the sake of quantity alone—is not
necessarily big thinking. Instead, it can be little more than a manipula-
tion of credentialing counts. In fact, years ago, when I was a graduate
assistant in the institutional review board (IRB) office, the IRB admin-
istrator told me that recruiting any more participants than needed to
answer a research question was a misuse of resources. Therefore, when
we think about big data in qualitative research, we should think not just
about big datasets, but also think bigger about what we can do as quali-
tative researchers—from envisioning projects that tackle bigger and
more socially significant research questions, to establishing stronger
academic–industry collaborations, to accessing hard-to-reach and
unique populations, to collecting and analyzing more complex data
instead of just more data, and so forth. These are the ways in which the
bigness of qualitative data in organizational communication research
gets really exciting!

Bisel: I so appreciate hearing you both say that. And, I agree. For me, a
large qualitative dataset is neither necessarily admirable nor problem-
atic. At times, I think researchers—and alas, often reviewers—may
have a certain faith in big numbers that probably comes from a genuine
concern about issues of generalizability and how generalizable claims
are supported by qualitative data. Of course, high-quality qualitative
research tends to be marked by sufficient immersion in the field and
with the interpretive worlds of participants. Sometimes, however, large
amounts of qualitative data are probably collected because researchers
lack a clear sense of design or what makes for a theoretically interesting
sample (Tracy, 2013). A large qualitative dataset is not necessarily indic-
ative of high quality. That being said, almost certainly there is a critical
lower boundary: Claims need to be supported; the strongest claims are
supported by a lot of evidence, even multiple bases of evidence.

I agree with Debbie’s point that a lot of big data work seems to revolve
around analyzing massive amounts of pre-collected data—collected
via social-media use or credit-card purchases, for instance. Yet in qual-
itative research, the researcher is often the research instrument and col-
lection is an integral part of the quality of claims that can be made. I
think part of the incompatibility between the big data movement and
the qualitative research in organizational communication is the lack of
emphasis the big data movement gives to the importance of data collec-
tion. The movement tends to presume great data are already collected
and waiting to be analyzed. Furthermore, collecting a large amount of qualitative data (e.g., conducting 100 interviews) could mean richness of observations were limited to achieve a large number of observations.

While somewhat rare, collecting a large sample might be warranted as a means of supporting claims of representativeness and generalizability, but that applies mainly for qualitative projects intended to be post-positivistic (e.g., Bisel & Arterburn, 2012). However, in much interpretive, qualitative work generalizability—in the post-positivistic sense—is not typically the goal. Instead, the goal tends to be the articulation of sensitizing concepts that capture contextually situated interpretive dynamics (Christians & Carey, 1989). Such sensitizing concepts can hopefully be transferable to similar contexts (Lincoln & Guba, 1985). For example, Gibson and Papa’s (2000) notion of organizational osmosis (a sensitizing concept) articulates an organizational socialization dynamic at a particular organization, in a particular blue-collar town and is not generalizable per se. Yet organizational osmosis is a powerful organizational influence on potential future members that we can see expressed in a number of ways, albeit, in contextually situated ways. Here, the point is not that organizational osmosis happens a lot or more than other kinds of socialization processes, but that it happens and how it happens.

Maintaining a systematic approach to analysis is another potential problem I believe arises with excessively large qualitative datasets. I participated in helping to analyze some very large qualitative datasets (e.g., Kelley & Bisel, 2014; Messersmith, Keyton, & Bisel, 2009; Minei & Bisel, 2013) and I know how incredibly mentally taxing it can be. The mind is a pattern builder and pattern recognizer but analyzing tens of thousands of lines of text stretches the boundaries of human mental capacity (see Kvale’s, 1996, chapter on the difficulty of interpreting 1,000 pages of transcripts) and may introduce a methodological weakness in its own right. I think this is no small point: Rigorous analysis needs to be systematic and exhaustive, but how is that rigor actually upheld where excessively large qualitative datasets are collected? I would venture to guess that often analysis is not systematic and exhaustive when the size of an unstructured qualitative dataset is excessive. Thus, from a researcher’s and reviewer’s perspective, I admire elegant design, collection, and analysis, which contribute to theory without needing to collect excessively large datasets.

Bisel: The big data movement has caught the attention of universities because it seems to have caught the attention of funding agencies. I am curious about your experiences with grant-funded qualitative research.
Barge: I agree that, in addition to traditional definitions of big data, we can think of “big data” in terms of the richness of data we collect within a limited amount of time in one site. When we think of big data this way, it highlights the importance of generating rich data. Let me provide an example: Over the past year, I worked on a National Science Foundation (NSF) grant to evaluate a communication simulation known as Prosperity Game that was used to generate ideas and possible solutions for broadening the participation of members of underrepresented groups in so-called STEM (science, technology, engineering, and mathematics) fields. Prosperity Games are multiday events where stakeholder teams, who are invested in the issue, are recruited to play (see Barge, Lee, Maddux, Nabring, & Townsend, 2008; Domenici & Littlejohn, 2007, for an overview of Prosperity Game methodology). In the present case, more than 40 people participated in 8 stakeholder teams with the purpose of generating recommendations that would broaden the participation of members of underrepresented groups in STEM. The simulation was crafted by a design team and conducted by trained facilitators. I classify our research efforts as big data, because the project utilized multiple data generation tools (interviews, surveys, observations, audio and video recordings, photographs, and worksheets) over time. It is an incredibly rich dataset.

I agree with many of the comments we heard so far and would add that the question of how much data need to be collected might be better framed as, “How rich do your data need to be in order to address your research question?” In our case, we needed to find a way to map the ecosystem of designers, facilitators, and game players, before, during, and after the game. On the contrary, if we were charged with answering the question, “How did game players perceive the game process and outcomes?” we might have generated much less material as this is a narrower question than, “What is the quality of the Prosperity Game’s process and outcomes?”

The impulse to go for big data is important to create a rich dataset that can be retrospectively mined. The “big data” metaphor allows us to enter a project being mindful of what data we need to generate to address our research question, but also being mindful of collecting rich descriptions of the unfolding communication processes and activities that we might subsequently return to and analyze. In the case of the Prosperity Game, this meant trying to document as much of the interaction as possible within and between teams, creating a rich dataset that we can engage with in different ways—some
planned from the onset of the project, some that emerged as we conducted the project, and some still unanticipated ways that are yet to emerge.

**Lucas:** Grants—even simply thinking about getting one—can enrich qualitative research. When I was a graduate student, I applied for a NSF doctoral dissertation improvement grant. The grant covered only direct expenses, so in developing the plan of research, I had to determine how I would spend US$7,500. The process opened me up to creative thinking about bigger possibilities. For my particular project, it meant I could recruit more participants (and a more diversified sample of participants) by running paid advertisements. I could interview participants in a broader geographic area by being able to fund travel to different cities. I could conduct my own primary archival research to frame the historical context instead of simply summarizing previously published accounts. Unfortunately, I did not get the grant. But the seeds of a bigger research plan were planted, making the original, smaller study pale in comparison. Motivated by bigger ideas, I found creative ways to fund (most of) the plan on my graduate assistant stipend. The point of sharing this story is that thinking about big money can encourage big thinking about research possibilities. Try it next time. Give yourself a number. What would you do if you had US$10,000, US$25,000, US$100,000, or more to execute a study? How could your project be enlarged to maximize your project’s contribution to theory? In addition to envisioning a bigger research plan, you may find that you have just boosted the fundability of your research.

**Tracy:** I have some experience with grant-funded qualitative research—most of it has been really valuable, but there are some cautionary tales to consider as well. Several years ago, I got connected with a funded research center that needed a qualitative expert—someone who could fulfill a large team-based project for which they had already contracted. They offered funding for one, then two, then three of my students. They provided funding and a teaching buyout. The actual work I would be doing remained fuzzy. I realized that the expertise of some of the most prolific grant-getters lies in making the deal and gathering a team of others. I was one of the members on that follow-through team.

This is when I began my travels in a place I call “evidence and big data land.” They were wandering travels because the expectations and tasks associated with the project were always evolving. They were occasionally worrying travels because I struggled with when to push for quality and rigor versus going along with what was asked by people who were not familiar with
qualitative research methods. They were wondering travels because it was new terrain, and as a curious ethnographer, traveling into this space was interesting and novel. What I learned from that experience taught me lessons about knowing when to lead versus follow, balancing the line between rigor or ethics and being disagreeable, and considering the trade-offs of doing what I thought was right versus going along for the ride.

For example, one of several tasks that our qualitative-guidelines group dealt with was dictating the level of de-identification of interview transcripts. Most qualitative researchers would agree that the meaning and value of qualitative data come from its rich contextuality—who said what and from which position and from what context. Often one cannot know what contextual clues will be interesting or important until after the analysis is underway. However, the majority of folks on this team instead felt as though a stringent de-identification needed to be completed to attend to IRB issues—and because we were sharing the data analysis tasks with each other, they felt the de-identification needed to be completed prior to the analysis. I tried over and over again to persuade the group that context was the key for intellectual rigor and linking our qualitative findings with the quantitative part of the study—and that it would be worth it to go back to our IRBs to seek permission to de-identify after the analysis. As I would learn later, my pleas were not successful.

As we progressed in our work, I realized that in this grant, the principal investigators (PIs) budgeted for the time needed to collect the data, but not to analyze or interpret it. This leads to the fact that my travels in this big data land were not just wandering, but also occasionally worrying. This worrying aspect of my travels would best be described by explaining the scene as all the PIs and key personnel were called to a central meeting, along with representatives from the funding agency, for a project update. As part of this meeting, the qualitative guidelines (including the de-identification issue) were going to be presented and voted on. Let me provide a picture of the scene of this 2-day gathering: The meeting begins at 8:00 a.m., sharp. My home university PI and I, jet-lagged, walk into the meeting room 15 minutes early, steaming coffee cups in hand. Three long tables are arranged like the letter “U” in the center, surrounded by about 20 chairs. A PowerPoint screen and lectern are set up front. Around the table are three additional groupings of chairs. Through my hazy sleepy-eyed fog, I begin to follow my PI contact—the only person I know in the room—to the table. I figure, we are there early, we get to sit at the table. Then, he looks at me and says uncomfortably, “Uh, Sarah, yeah, this is set up in their traditional style.” I look at him having no idea what he means. He seems embarrassed but continues, “Uh, that means that only the PIs sit at the head table, and the rest of folks sit in the, the,
peanut gallery.” The peanut gallery? Did he just say, “peanut gallery”? Yes. Apparently, that is the name for where all the non-PIs sit. Fast forward a day and a half through the meeting. During this time, various PIs rose from the main U-table to provide presentations. Then, we get to the qualitative guidelines. The PI who led our conference calls rises to the front of the room and announces that she is not a qualitative expert, but because no one else stepped up, she led the creation of the qualitative cross-protocol guidelines. Instead of presenting the guidelines herself, they are presented by other members on our team who helped devise them. Given my disagreement with many of the mandates, I opted not to present them.

I watch after one, then a second, then a third, then a fourth, then a fifth member get up to present the guidelines. Each of them walks carefully, shimmerying themselves through the rows, and finding a place to set their coffee cups, as they emerge from—drum roll—the peanut gallery. Whereas the previous speakers were seated at the central PI U-table, those talking about the qualitative arm of the research are seated on the periphery. The scene is richly metaphoric, and I cannot help but wonder whether Goffman (1961) would conceptualize this entire performance as a degradation ceremony. After the guidelines were presented, I provided a few comments, gently promoting alternative ways to proceed. I am nervous, out of my element, but try to speak as confidently as one can—from the peanut gallery.

Finally, a vote is called whether or not to approve the qualitative cross-protocol mandates for all the research centers. Only the people in the inner U got to vote—all the people who say they know little about qualitative research. They are the deciders. All those approve say “aye”; a resounding aye goes forth. All those opposed? One lone hand wavers and then raises—the PI from my home university.

All this says a bit about my worrying—worrying about whether I should have just gone along; worrying about whether I should have taken a bigger role and tried to lead the group elsewhere; worrying about the eventual value of big data–funded projects when subjective qualitative data are stripped of their contextuality and when the focus is on the logistics of compiling data rather than analyzing their meaning.

However, my travels in this big data evidence land also included some wonder and wondering. As a qualitative scholar, I was curious about interdisciplinary multiple-site funded research. Grant work has a revered quality at my home institution, and I was curious to dip my toes into the grant pond. This project provided firsthand knowledge of both the shiny and not-so-shiny bits. Here, I recommend Cheek’s (2011) chapter on grant-funded qualitative research. Like me, she also reflects on motives: Is funded research about “enabling research or gaining the funding” (p. 253)?
Some “wonder”—ful parts of the experience include that I felt like qualitative expertise was needed (if not always appreciated or understood). I found myself thrust in the midst of those who admittedly knew next to nothing about qualitative research and asked for help. Methods that are familiar to me—like qualitative codebooks, intercoder reliability practices, and iterative data analysis—were welcomed as brand-new sets of tools and expertise. Another wonderful part of the project was the ability to be a job creator. Through the project, my work on the grant facilitated research assistantships over eight different terms to four different students, and this allowed some talented students to practice research methods in an interdisciplinary environment. However, given my limited status on the team, the publication prospects—both for me and my students—are still unknown. Another “wonderful” part is the access. As an organizational ethnographer, qualitative access to organizations is always a struggle. By being part of a grant team that negotiated long-term relationships with various organizations, I was instantly provided access to data in a range of institutions that otherwise would have been off-limits.

**Bisel:** Thanks for sharing those experiences, Kevin, Kristen, and Sarah. I am interested in discussing what you believe should be the role of qualitative research design when working with big data.

**Lucas:** Qualitative data collection can be an exhilarating experience, but it is extremely time-consuming. Because of how much energy is invested in gathering a big qualitative dataset, it is helpful to think about overall research goals, how the current study is positioned in that line of research, and how it might be strategically linked to the next project. One of the practices I started is asking the “next question” while in the field. By building the next question of my research agenda into my research design for a current study, I maximize my data collection efforts. For instance, when I conducted my dissertation research on career transitions in a postindustrial economy, I knew that the next step in my research agenda was going to be a shift to the topic of workplace dignity. So, while I was interviewing my 60 dissertation participants, I asked one more question about dignity with a couple follow-up probes. Then, I ignored those answers as I performed my dissertation analysis. When I was ready to move forward on my research agenda, I had a “brand-new” set of data without having to return to the field (see Lucas, 2011b). It was quite the time- and money-saver.

**Barge:** Qualitative research design needs to be given even greater attention when dealing with big data. Consider a typical qualitative organizational communication study: Researchers negotiate access to an
organization or organizations, which permits follow-up interviews and other kinds of ongoing meetings to check out the analysis. The design of the study and subsequent analyses are often emergent because of the way the researcher negotiated ongoing access. Now consider how these assumptions may change with big data. We may now have several hundred interviews or multiple sites where data are collected. This means the researcher may not be the person who conducts the interview or observation. This may even take the form of standardizing data collection among multiple sites, hiring other researchers to collect data, and in some instances, being the recipient of a dataset that the researcher did not personally design or collect. In each of these circumstances, this means researchers may not have the opportunity to go back and collect additional data or member check their analysis. Under these conditions, there is more pressure to (a) be mindful of how you structure the design of your study to make sure that you generate the data you need as you may not have subsequent opportunities to collect additional data and (b) in the case of working with a dataset that is provided to you, to think creatively how you can analyze the data and use different pieces of existing data to cross-check your analysis.

In the case of the Prosperity Game (described above), we knew we would not be able to go back and re-interview members of the design team and facilitators. We also knew that we would have one shot to capture the dynamics of the game. We had to plan carefully how to capture the live game play as it unfolded. This required us to think hard about design. Eventually, we decided to document, in real time, using audio and video recording along with photographs to archive the unfolding interaction as closely as possible. I would label the strategy we used for data collection as planned improvisation. Simply, we knew that there were certain kinds of research questions that we wished to pose. However, we also wanted to build in the opportunity for generating data that may be of theoretical interest later. Similar to Silverman’s (2010) notion of progressive focusing, we fine-tuned areas of exploration as the game and data collection unfolded. But this occurred in a framework of being clear about what data we needed to collect initially and how we could collect them, which is the essence of design.

Dougherty: I agree, Kevin. For me, most of this discussion comes back to research design. Specifically, research design is how we determine when big qualitative datasets are necessary in the first place. The answer is both simple and complex. The size of the dataset should be fully dependent on the research questions and the resulting study
design. In other words, the value of big qualitative data depends on study design. Qualitative research can be elegant in its simplicity, with high-quality interviews that help reveal important communication processes. Each of you on this panel has excellent examples of this type of research. Your studies are well designed and crafted; they are systematic and really quite lovely as both art and science.

Bisel: I appreciate those points. Design is about helping yourself get to theory building—the goal of inductive research. Theory building seems to happen when analysis can be situated in such a way as to “speak to” a body of knowledge that makes others take notice. I noticed that getting to the point of building theory is not merely about solving puzzles, but rather, finding puzzles I can solve that are worth solving. In the design stage, I ask myself, “Where can I find data that might reveal how common sense and theory about this topic are mistaken or can be extended?” My graduate students directed me to a fascinating study by Davis (1971). In the study, Davis provides a phenomenology of “interesting” social science theories and reveals how interesting theories disconfirm assumptions held by the audience and, therefore, leaves the audience with an emotionally charged sense of “ah-ha!” Clever design can help us get to readers’ ah-ha’s. My graduate students and I now say to one another, “Did you Davis it?” to imply the need to keep refocusing on where data can help us contribute to theory. Big qualitative data are “big” because they help us extend theory, not because they are so numerous. Yet, I admit, it is quite easy to fall into the trap of thinking that more is more. The better I am able to keep my focus on the goal of theory building (and synthesis with the literature; see Lucas & D’Enbeau, 2013) throughout the process, the better I am able to design research, locate theoretically provocative samples, and collect data that are interesting and newsworthy.

Tracy: Well said. I saw firsthand how qualitative design can be shaped by big data grant funding. Perhaps the most remarkable admission about my work on the granted research (described above) is that it was not until a couple years into the project that I figured out the exact foci of the study. Engaging with a research project after it was already designed and conceptualized was completely new to me. In my own “small” qualitative studies, I was always intensely interested in and motivated by a specific issue or problem—something I call a contextual, problem-based, phronetic approach (Tracy, 2013). This granted project was so huge and multifaceted that it was difficult to understand the main issues or problems that needed qualitative research attention. I wondered, “How is rich qualitative design even
possible in such a situation?” My typical approach is to start with a problem (e.g., burnout among correctional officers, see Tracy, 2005, or workplace bullying, see Tracy, Lutgen-Sandvik, & Alberts, 2006) and then engage in qualitative investigation that tries to shed light on the issue and provide a space for transformation. Along the way, it became clear that a primary part of my role was to help set qualitative guidelines for gathering, conducting, and analyzing data across a number of research settings. In past projects, most meetings with collaborators focused on research ideas and intellectual areas of inquiry. In contrast, this big data qualitative granted project required many discussions about logistics and coordination. It is good that most of these meetings were conference calls. I found practicing yoga at the same time kept me calm and present.

After several months into being involved on the calls, I learned that the person leading the qualitative-guideline group knew next to nothing about qualitative methods. Why was she heading the calls? Because they needed to have a PI lead our group, and none of the PIs (of about 10) were qualitative experts. They involved qualitative folks (like me) to carry out the study, but not to design or lead it. In my particular case, I was not brought on until halfway through the study’s completion, so I was an outsider even among the non-PI qualitative personnel. Furthermore, it was an interdisciplinary group and I was the only communication scholar involved. As such, I entered the project with absolutely no history or established credibility. I faced a dialectic between wanting to speak up and encourage best qualitative practices, but at the same time, realizing I was new to the scene, sitting back to learn how such projects unfolded.

**Bisel:** Those comments really illustrate the complexity surrounding design in big qualitative data. Thank you. What important tips can you share about qualitative data collection with regard to big data?

**Lucas:** From the outset of a project, it is important to think carefully about just how big a data collection will become and how its size will affect the planning and execution of the study. One of the exercises I recommend is a basic computation of time involvement. Instead of making wild estimates of time (e.g., “I will collect data for a month”), the exercise involves breaking major components down to their most easily estimable parts and then calculating. With an interview study, for instance, I project how many interviews will be conducted and how long I anticipate them to last. I then add travel time to and from the interview (1 hour per interview), processing time (1 hour per
interview; copying audio files, filing IRB paperwork, writing a short recap memo), and transcription time (4 hours per hour of talk). Then I divide the total by the number of hours of work I plan per week. For example, thirty 1-hour interviews (plus the additional 6 hours for travel, processing, and transcription for each) require 210 hours of work. If I have 20 hours per week to dedicate to research, data collection takes 10+ weeks. I do the same for archival analysis. If I am researching coverage in a daily newspaper and estimate 3 minutes per issue for scanning the headlines, and downloading and printing relevant articles, that culminates in 18 hours (or three 6-hour days) in front of a microfilm machine for each year of coverage. When I began doing these computations, I realized that I was grossly underestimating the amount of time involved in data collection. These estimates allowed me to focus my data collection efforts on a smaller and more targeted area of interest, plan for extended time in the field (especially when it involves travel), and adjust weekly time commitments to finish by a deadline.

Another absolutely essential task to doing a big data collection is having a data management plan in place before collecting the first bit of data. Managing a large dataset gets overwhelming quickly. Therefore, I recommend developing a plan for sorting and cataloging it all prior to collection. Each data collection will have different needs, but here are a few pointers that worked for me: If you are interviewing people, that likely means having several pieces of information on each person (e.g., recruitment communication, informed consent documents, information sheets, handwritten interview notes, audio recordings, transcripts, interview memos). Moreover, some of these items are going to be hardcopies and some are going to be electronic. Knowing that this task is going to be further complicated by the addition of pseudonyms, I start with a numbering system and then create a paper file folder and an electronic file folder by the same code. I keep all my paper files in a large portable file box that I can easily carry to interview sites (for immediate filing), and I keep all my electronic files in a master folder on my computer, which is duplicated and saved on an external hard drive.

Along with the data management plan, I always keep a summary spreadsheet for referencing the big picture of the data collection. In the case of tracking participants, this includes summary information like participant number, pseudonym, age, sex, occupation, job title, or any other characteristic relevant to the study. During the data collection process, I can monitor the spreadsheet to make sure I am getting a well-balanced mix of participants so I can make adjustments to recruiting efforts, if necessary.
Dougherty: When it comes to data collection, tip number one is “No short-cuts.” Collecting and analyzing big qualitative data takes time and attention to detail. Many quantitative scholars can conduct multiple quantitative studies in a semester. This is not likely to work for a qualitative scholar. Collecting qualitative research takes a lot of time, so relax and go with it. Collecting a big qualitative dataset is a huge time commitment. It may take a full year of concentrated attention to collect and analyze the data. Tip number two is “Do not wing it.” Given the amount of time commitment necessary for big qualitative research, it is important that every study yields usable results. My own experience has taught me to make sure all research personnel are well trained in the necessary methods. It can be devastating to realize, as you read through interview transcripts, that the interviewer did not know how to probe and therefore missed most of the important cues that warrant further investigation. I also had interviewers skip questions because they felt the issue had been covered when, in fact, they were not addressed. Train yourself and your people so you have the skillsets to improvise. Yes, qualitative interviewing is kind of like a conversation, but also kind of not like a conversation. At the end of an interview, if you are not tired from listening deeply, then you may not have been fully engaged.

Barge: My experience with the Prosperity Game required that I deal with a combination of a large dataset and the necessity of collecting it within a naturally occurring event. That experience taught me three key lessons: First, we felt it was important to preserve the sequence of the unfolding event and sample repeatedly over time. As a result, we designed several moments before, during, and after the game to generate data. By sampling data at different points in time, our hope was to see how patterns of communication would change over time. Second, we needed to collect unobtrusive data measures. In Webb, Campbell, Schwartz, and Sechrest’s (1999) classic book, Unobtrusive Measures, they highlight several ways to generate data that do not influence participants directly and minimize interference with their normal activity. For example, we were interested in mapping the social networks of game players throughout the game. Rather than have players complete written surveys at different moments, we used existing worksheets that the teams completed where they listed the teams they talked to at two different points in time. We also took photographs of the entire game space at several points in time to map the networks. This allowed us to use both subjective and objective data to map the networks. I also recommend considering video recording as a great means of collecting a lot of qualitative data that can be revisited.
Third, it is important to have a research-design team with people from different backgrounds. One idea that I take seriously from the literature on engaged scholarship and research (Van de Ven, 2007) is the notion that difference is a valued resource. Creating teams of researchers who have varied experiences leads them to engage the same problem or material differently, bringing several different perspectives to bear on the issue at hand, which generates a richer set of ideas. For example, in the Prosperity Game, my primary co-researcher was trained in quantitative methods. He brought a different set of tools and experiences to the design of the research than I had and made it richer. Similarly, we reached out to several scholars in the communication and STEM disciplines for ideas regarding the kind of data and analysis we could use. One colleague suggested use of photography as a way to collect network data—an incredibly simple and elegant approach we had not considered. Another colleague suggested that we record the plenary sessions and also collect flip charts so that we could do linguistic analyses on the material and see whether the language of various teams changed or remained the same over time. The point is that we drew inspiration in the design and analysis stages by including people with different research experiences than our own.

Bisel: Those are practical tips. How does a big qualitative dataset shape your approach to data analysis? What tips can you offer to a qualitative researcher who is having difficulty analyzing a large dataset?

Dougherty: Qualitative data analysis is amazing, fun, energizing, hard, and tedious. I love the discovery part of this process. However, if you are not prepared to spend time in your chair doing the tedious work of qualitative analysis, this may not be the right research approach for you. Complex qualitative designs that generate big data are even more difficult and tedious. I once spent an entire summer analyzing a dataset (Dougherty, 2001, 2006). I also once spent 40 hours over a 2-week period designing and conducting a divergence analysis of a dataset (Dougherty, Kramer, Klatzke, & Rogers, 2009). I do not enjoy sitting for long periods, but it is an amazing experience when knowledge finally emerges. My advice is to design an analytical process that will answer your research questions. Then put your butt in your chair and push through with that process. I think of qualitative research as akin to the great explorers of the past. Sacagawea, Lewis, and Clark undoubtedly did not always have a grand time in their exploration. They got hot, sweaty, bored, and fatigued during their journeys. The discovery is worth the hardship. If you want to create something amazing then suck it up and do the work of being amazing.
Lucas: Ryan mentioned earlier that analyzing big datasets can stretch the boundaries of human mental capacity as we try to recognize and build patterns. That is so true. When a dataset gets really big, our brains simply do not have the capacity to keep track of everything we have heard or read. Personally, I tend to remember the amazing quotations or particular interviewees who make strong impressions, but I still forget many good quotations. That is why it is important to use tools like qualitative data analysis software to compensate for mental limitations. By coding with qualitative data analysis software, I can retrieve a report on a particular theme and then dig into the complete set of relevant excerpts without missing any that I forgot. Qualitative data analysis software also allows for multiple coding of the same passage, renaming and refining codes without having to recode manually, complex retrievals, and more.

Another approach that I recommend for working through the difficulties of analyzing a large dataset is holding early intervention data sessions (see Lucas & D’Enbeau, 2013, for details and additional analysis ideas). In this kind of session, I invite key individuals (usually peers with theoretical or methodological expertise) to an interactive workshop-style presentation. In the session, I identify the research question, provide a brief description of the data collection procedures, and then spend the majority of the time presenting preliminary categories or themes, including conceptual definitions and key exemplars. I encourage attendees to ask tough questions throughout the presentation about why I thematized data in a particular way and push me for alternative interpretations. By doing this early in the process of analysis, instead of near the end, it can help me confront complications early and gain big-picture insights about the data. And it is so much easier to make the necessary changes before particular understandings get too entrenched in my thinking and before I committed too much time and energy to pursuing ideas that will not be fruitful.

Barge: I learned that our research team benefitted from developing simple analytical tools for engaging the large amount of material. Simple analytic tools may be useful to orient yourself and identify the foci for further inquiry. In the Prosperity Game research, we often began our analysis with simple content-analytic schemes or frequency counts to help focus our analysis. For example, we were interested in how the recommendations of the various stakeholder teams changed over time. We developed a simple content-analytic scheme using three strategic intents to classify the recommendations for broadening participation in
STEM. The scheme helped us characterize and understand how participants’ use of language reflected issues of infrastructure, engagement, and research. This provided us with a simple way to map the ebb and flow of recommendations. We found that they did shift over time and now we are in the process of determining how we can conduct more nuanced linguistic analyses of the content so we can map linguistic shifts over time and between teams. But we would not necessarily have pursued this line of inquiry if we had not noticed a difference in the frequency of different types of recommendations over time.

**Bisel:** Those are great points. My original inspiration for assembling this conversation came from one of my experiences as a dissertation advisor. A couple years ago, my advisee collected 40 very lengthy and intensive interviews with leaders who had reputations for being especially skillful. The dataset rendered nearly 1,000 pages of transcripts—just picture a genuine line-by-line analysis of more than 20,000 lines of text! Eventually, we admitted to ourselves the dataset was simply too large to approach with a conventional constant comparative analysis because open coding was generating an unwieldy number of codes. Deliberate, systematic, and exhaustive coding of the dataset was an unrealistic task for one mind, at least all in one chunk. Eventually, my advisee and I devised a plan and discovered that two strategies were especially helpful for her: First, *data reduction*—the retaining of data only needed for answering the research questions—was important for focusing analytic attention. That being said, my advisee needed to explore the data enough in terms of early and tentative coding to know exactly what research questions could be asked in the first place. In that sense, this process was iterative and required her to act first to think (Weick, 1995). Defining carefully what should get coded to answer research questions (i.e., data reduction) is a process that probably tends to be accomplished during open coding for most researchers; however, such reductions are not usually described explicitly in Method sections. Explicit explanation of data reduction processes may be especially important and useful when a large amount of unstructured qualitative data needs to be analyzed.

Second, and relatedly, we devised a plan in which she began with 5 of the 40 interviews, treating those 5 as the dataset. She conducted data reduction, open coding, and categorical consolidation—applying the contemporary conventions of constant comparative analysis in communication research (Lindlof & Taylor, 2011). She repeated the entire process completely with another 18 interviews, but this time she also paid special attention to
disconfirming, amending, or adding to categories determined in the first stage, allowing her to compare and then synthesize the two analyses. The final 15 interviews were then used to see if the categories rendered from the preceding stages could account for the remaining data comprehensively. The process worked well for cross-checking the analysis and provided a platform for pattern recognition in digestible chunks (see Kelley & Bisel, 2014). On deeper reflection, I wonder how often qualitative organizational communication researchers, who collect very large datasets, engage in such strategies? Also, I wonder if such strategies are more commonplace than get reported. There is likely no way of knowing, but I think being able to articulate modifications to conventional analytic practices is crucial so that we can learn from and continue to explore techniques for grappling with tricky analyses, especially those involving big qualitative data.

**Bisel:** *Is there anything else you would like to offer?*

**Lucas:** My advice on doing qualitative research is one I learned the hard way: Tell the largest story possible—even if it means fewer published manuscripts. When I got done with my dissertation, I wanted to publish several articles from the data. I started by parsing my dissertation into the smallest pieces I thought were publishable. Then, I went through a few devastating years of journal rejections. At the most abstract level, the reviews always seemed to be the same: There was something missing from the analysis. I would throw my hands up in frustration and think to myself, “But it is in the other manuscript!” Eventually, I changed my strategy and decided to tell the biggest story possible with my data. I challenged myself to the daunting task of refining a 100-page section of my dissertation until it could fit into a journal-length manuscript. It was not until I stitched the pieces back together into a larger whole that I started having success publishing (see Lucas, 2011a, 2011c). Moreover, my advice to tell the biggest story is more than just a strategy for getting published. It also is a commitment to communicating an important message about qualitative research and what it has to offer the field of organizational communication. Qualitative research—and its bigness—is not found only in its numbers but in its potential to build theory.

**Tracy:** I would like to add that one of the wonderful aspects of my experience on the big data granted project is that the impact of the work is bigger than me. Although I sometimes felt like I did not have a handle on the entirety of the project, this also meant that things bigger than myself emerged from it. In one case, for instance, our data analysis served as a basis for promotional materials. Right on the brochures
were phrases that we identified in our qualitative data analysis—comments like, “With medicated assisted treatment, I now have money in my pocket.” Working in a large granted group provided the skills, abilities, and means to move our findings into materials available to the affected populations. As such, it was wonderful to see the larger effect (outside of the scholarly publishing we were conducting) of our qualitative analysis efforts on this project (Malvini Redden, Tracy, & Shafer, 2013). There is no conclusion to this story. I continue to learn, grow, and try to understand the advantages and disadvantages of being involved in large qualitative funded projects. Hopefully, my cautionary tale can help inform others as they consider the opportunities and restraints of practicing qualitative research in big data complex research venues.

**Barge:** The driving question for me is how to keep one’s sense of reflexivity alive during all stages of the research process. A good deal has been written about reflexive research practice (e.g., Alvesson & Skoldberg, 2009; McNamee & Hosking, 2012). The intent is to find ways to challenge your thinking and avoid falling in love with your own hypotheses. I think this is particularly important when we consider working with big data. How do we simultaneously design research that has a guiding intent while remaining open to new puzzles, challenges, and dilemmas our evolving analysis suggests? Big data require us to be curious about the data and sensitive to emerging possibilities. My suggestion is to invite disruptions deliberately and provoke your own thinking. I recommend inviting disruption by (a) reading widely—both inside and outside the communication discipline—fiction and nonfiction, (b) engaging in conversations with a variety of academics and practitioners, and (c) asking yourself “What if?” The challenge is to keep curiosity alive and retain a sense of playfulness and irreverence when designing your study and analyzing your data so that “big” (read, theoretically interesting) questions can be answered.

**Conclusion**

**Bisel:** Our conversations at conference and here in the pages of *Management Communication Quarterly* cause me to reflect on how we are answering the question, “What does the big data trend mean for qualitative organizational communication research?” It seems that each of us—in our own way—is emphasizing the need to think carefully about the meaning of “big” to include depth, richness, and theoretical importance. We are suggesting that the “bigness” of data should be
understood in relationship to the questions that can be answered and the theory that can be extended.

Acknowledgments
Ryan S. Bisel would like to thank Josh Watson and Dr. Katherine M. Kelley for their role in inspiring this manuscript, Elissa N. Arterburn for chairing the 2013 conference panel, Alaina Zanin for providing important feedback during the manuscript’s development, and Dr. Ling Chen for her efforts in making this manuscript a reality.

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

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

Note
1. Dougherty: I know the term “Triangulation” is currently out of vogue with some qualitative scholars because it suggests a reality out there. However, I use this term in the larger sense of more perspectives.

References


Author Biographies

**Ryan S. Bisel**, PhD, University of Kansas, is an associate professor in the Department of Communication at the University of Oklahoma, United States. His main research interests include leadership communication, organizational culture, and behavioral ethics.
J. Kevin Barge, PhD, University of Kansas, is a professor and head in the Department of Communication at Texas A&M University, United States. His major research interests center on developing a social constructionist approach to leadership, articulating the connections between appreciative practice and organizational change, as well as exploring the relationship between discourse and public deliberation, specifically practices that facilitate communities working through polarized and polarizing issues.

Debbie S. Dougherty, PhD, University of Nebraska, Lincoln, is a professor of organizational communication at the University of Missouri, United States. Her research focuses on organizational power, particularly as it relates to emotions, social class, and gender.

Kristen Lucas, PhD, Purdue University, is an associate professor in the management department at the University of Louisville, United States. Her main research interests center on workplace dignity and career discourses.

Sarah J. Tracy, PhD, University of Colorado, Boulder, is professor and director of The Project for Wellness and Work-Life in the Hugh Downs School of Human Communication at Arizona State University, United States. Her research focuses on emotion, work–life wellness, and qualitative research methods.