

The Fieldwork of Quantitative Data Collection

Francesca Refsum Jensenius, *Norwegian Institute of International Affairs*

FOR many political scientists, fieldwork means conducting focus groups in villages, attending campaign rallies, or interviewing political elites in government offices. When I present the mainly quantitative findings from my PhD work on electoral quotas for the Scheduled Castes (the former “untouchables”) in India, colleagues are sometimes surprised to hear that I spent more than a year conducting fieldwork for the project. To study the effects of quotas in India I wanted to combine statistical work with interview-based case studies. I collected some of the quantitative data needed for the project during two initial field trips, and then returned to India for another nine months of fieldwork, intending to conduct interviews and collect more data. The main surprise was how easy it proved to get interviews, whereas considerable time and effort were needed to get access to “publicly available data.” This article is about some of the failures and successes of my fieldwork, focusing particularly on the social relational aspects of collecting quantitative data.

Fieldwork-based work is often contrasted with quantitative data work, but while some quantitative datasets can be downloaded from the Internet or bought from data-collection agencies, other datasets are the result of months and months of pestering officials, searching through archives, or accompanying data-entry people in the field. To gather this type of data, one spends considerable time on both gaining access and building rapport with gatekeepers, topics familiar from discussions of qualitative data collection (Berg 2003; Brooke Harrington 2003; Scoggins this symposium). Local knowledge also gives insights into how large datasets are collected, where their weaknesses lie, and how to spot irregularities in the data. This insight can be key to ensuring data reliability, an issue frequently discussed in methodological texts for political science (e.g., Kellstedt and Whitten 2013, chapter 5). By sharing some examples from my own field trips, I hope to show the importance of fieldwork for quantitative data collection and ways of dealing with the frustrations resulting from trying to collect data in the field.

GETTING INTO THE BUILDING

The first hurdle in trying to access quantitative data is how to gain access to the building where the data are stored. This is a very physical and concrete version of ethnographers’ challenge of “entry” (e.g., Johnson 1975, 52). My fieldwork was full of frustrations related to getting into buildings, compounds, and archives. In India’s largest state, Uttar Pradesh (UP), the legislative assembly and its archives are surrounded by a tall fence with intimidating gates and guards. When I first came there to consult the archives, I was pointed to a small office by the entrance gate that issued entry passes. There was a long line of people waiting, and

since I did not want to use my “foreigner-card” to skip the line, I waited there for a long time. When I finally came to the head of the line, I was told that I could not get an entry pass unless I had an appointment with someone working inside the compound. The legislative archives in India are supposed to be open to researchers, but although I showed my research visa and letters confirming my academic affiliation I was told I could not get access without such an appointment. Because I had previously visited archives in other states of India, I insisted that I was entitled to access the archives. The officer on duty then told me that I would need permission from the head librarian, but when I called the head librarian to ask to see her she told me I could not meet with her unless I had an entry pass to enter the compound! I finally accepted defeat. Fortunately, a colleague with whom I was traveling had some local political connections who arranged an appointment for me with one of the head civil servants working inside the compound. Once inside, it was easy to get the additional permissions.

A similar situation occurred in the state assembly of Haryana, also in northern India. In this case my colleague and I passed through the main gate of the legislative assembly compound by showing letters proving our research affiliation and explaining that we wanted to access the archives. Here the challenge was to get into the actual archives because the staff at the reception desk claimed that only politicians were allowed to enter. Here too we insisted, and in this case I believe it was the fact that I as a foreign woman pleaded to them in Hindi that made them soften up and allow us access. The staff were not following any procedure: they made an arbitrary choice of granting us access.

What these stories show is that to get through the doors where data are stored you often need to be persistent, use contacts, and plead nicely to gatekeepers for access. This can be frustrating, humiliating, and time-consuming. For me it has proved to be a huge advantage to travel with a friend or colleague, and I now try to do that as often as I can.

CONVINCING GATEKEEPERS TO GIVE YOU DATA

When inside the right building, the next step is to convince the people who have access to the data that they should give it to you. This too can be time-consuming, and is often about building trust and “rapport” in much the same way as researchers who collect qualitative data (see Glesne and Peshkin 1992; Marcus 1997; Scoggins this symposium).

In one case I was trying to obtain some publicly available education data in UP. During my fieldwork in the northern state Himachal Pradesh (HP) I had discovered a fascinating survey of infrastructure, teachers, and students covering all public schools in India. The civil servant in charge of the data collection had

given me the entire raw dataset for that state, so I was excited about collecting it for the state of UP, too.

The office responsible for collecting the education data in UP was about half an hour's travel from where I was staying, and I went there in an auto-rickshaw with an Indian colleague. When we got to the correct office we asked for the data-entry people. From previous experience, I had learned that it is vital to know exactly what is available on file before making requests to the officials in charge. Having ascertained that they had all the data I wanted, we then asked whom we should ask for permission to get the data. We were sent to one civil servant, but because he was away for the day we were told to come back the next day. Not too disappointed, we traveled the half-hour back to our lodgings and came back again the next morning. When I met with the civil servant and explained what data I was looking for he told me that nobody had ever asked for this data before and that he was not sure whether he could authorize giving it to me, so he sent me to a higher-level official. However, that person was not in the office, and I was told to come back another day.

In this case I did not get the data I needed because I failed to create the feeling of trust necessary for them to believe that my intentions were really to do long-term research and not to create a media scandal.

The following day I was sent to yet another person, the head of the department. He told me that I needed to submit a written application for him to consider my request. I left his office, wrote up an application on my laptop, printed it in a shop down the street, and returned to his office with the completed application. By that time he had gone out for lunch. So we waited for him for two hours, but he did not return. By now a bit tired of the situation, we traveled back to the city center again and returned the next day to give him my application. He told me to leave the application with him and that he would get in touch with me. However, wise from earlier experience, I insisted on waiting. He then told me to wait outside his office, but then left through another door and did not return for many hours.

Realizing at this point that the officials simply did not want to give me the data, I asked my Indian colleague to make some enquiries. From the civil servant we had first approached, we learned that there had been a lot of internal discussion about my visits, and that the leadership had decided not to give me the data because they were worried that I would discover the poor quality of the data. Apparently, there had been major problems with how the data had been collected and coded, and if we studied it we might discover some of these weaknesses. Nobody wanted to take the blame for having given me the data in case I should publish something that resulted in a public scandal. In this case I did not get the data I needed because I failed to create the feeling of trust necessary for them to believe that my intentions were really to do long-term research and not to create a media scandal.

In the end I accessed this data through the central office in New Delhi. However, having learned about the poor quality of the data from the UP office, I was far less enthusiastic about it than I had been initially. This story shows the importance of building rapport and trust, as well as the importance of trying

several avenues for getting the same data. It also shows the value of traveling with a colleague, in this case a local scholar. Being a foreigner I can usually gain access to high-level officials, but when it comes to hearing about office gossip, being local is a huge advantage. When traveling back and forth to the office, and then waiting for hours, it is also nice to have company.

As the previous example shows, trust is central to getting access to data. In two other cases I was initially refused access to data sources because other scholars had broken the trust of people working with the data. In one case I was refused access to an archive because another scholar had taken pictures of data sources although this was explicitly not allowed. The librarian was upset about this disrespectful behavior and took out his anger on me. I was consequently refused access to the documents I needed, although the person in charge of the archive had granted me access. It took several hours of drinking tea with the librarian to calm him down and convince him that I would indeed follow the rules.

In another archive I was refused access because another foreign scholar had tried to get some data from an archive, and, finding

the process too slow, had gotten a powerful political friend to put pressure on the librarian. The librarian was deeply offended by this behavior, and because I was the next foreign scholar to come along, she gave me a long speech about how disrespectful all foreigners are and how it gave her a "bitter taste in the mouth" to help us out with our work and then get this kind of behavior in return. She consequently refused to help me, and again I had to spend considerable time talking with her about my work to gain her trust and be allowed access to the resources in the archive.

These experiences were frustrating, but also taught me the important lesson of always being respectful and polite to all the officials I encounter in my work. The importance of being respectful is often discussed in connection with qualitative fieldwork, but not in the context of quantitative data collection. The people in charge of entering, storing, and administering data are often hard-working individuals who do not receive much gratitude for the work they do. They must take time out of their already busy schedules to help researchers who come to request data. Naturally, they feel upset when they find that their work, time, and operating procedures are not respected. Being respectful, as well as patient and persistent, has therefore become a major rule for how I approach data collection.

DATA LOST, DAMAGED IN A FIRE, NEVER EXISTED

Another challenge arises when those who are supposed to have data claim that the data never existed or cannot be found. This is often not out of ill will, which means that neither good access nor good rapport is helpful. A clear example of this occurred during an early field trip, when I was trying to obtain lists of villages that fell under each political district in India to merge political data with development data. Expecting this information to be fairly readily available, I went to the Election Commission of India to ask for

it. And here I believe the official I talked to was willing to provide the data—but when he asked one of the men working in his office to give me the CDs with lists of villages for each state, they could not find them. They were sent out to search, and came back with several CDs that I was allowed to view on my computer and copy. Some of them contained information I had been looking for, but data for several large states were missing. After many rounds of phone calls we heard that the remaining CDs had been lost in a flood in one of the basement offices during the last monsoon. I do not know whether these files were actually lost in a flood, or had simply been misplaced. What I do know is that a few years later, when I returned to the same office to ask for some other information, the officer in charge asked me timidly whether I would be willing to share with them the data for the states I had copied, because they had misplaced more of the CDs and no longer had access to this information themselves.

Here is a final challenge: although some data may be easy to obtain, they may prove unreliable or useless.

In another case I was trying to get access to the district-wise census booklets that the Census of India had prepared for the Election Commission of India for use in delimiting new political districts in the 1970s. After a few visits to the Census office and the Election Commission offices, where I was varyingly told that these documents had never existed or had never been archived, I was finally sent to an obscure archive on the outskirts of the city, where copies of these documents were supposed to be kept. There I was told that the collection had been lost in a fire 10 years earlier. Later, I discovered these booklets in the Election Commission archives. Because these were historical documents that the Election Commission no longer needed, no one knew that they were there. This experience taught me to always look for myself, rather than simply accepting that something does not exist.

DATA RELIABILITY AND USEFULNESS

Here is a final challenge: although some data may be easy to obtain, they may prove unreliable or useless. Previously I discussed poor quality of the education data from UP. My visit to the HP archives to gather information about the bills introduced and passed in the Assembly over the years is another example. This information is available in the minutes of the debates for each legislative session that are stored in the archives and in booklets summarizing each of the debates. When I told the librarian what I was looking for, she enthusiastically explained that my work would be easy because one of the staff in the library had already gathered all of the information. And indeed, my colleague and I were soon handed a complete list of all the information we were looking for. Somewhat surprised at achieving our goal so easily, we asked whether we could still see the archives and the books. We soon discovered that many of the collated figures were incorrect. I do not know whether the person working on this had been sloppy or whether the information was gathered only from certain publications, for example only those issued in Hindi, but we ended up spending several days assembling a new version of the dataset, with quite different figures.

I will end with the story of one of my major disappointments in the field. After I established contacts in the secretariat in one Indian

state, a high-level civil servant promised to use his power to help me get data on how state-level politicians spend their development funds—a discretionary cash fund that politicians can allocate to development projects of their choosing within their political districts. He told me that the secretariat kept records of the spending of the funds and usually did not share this information, but that he would do me the favor of having it entered for me in Excel format. Having high hopes for the usefulness of this data, I returned to that office many days in a row to follow the progress of the data entry. After several days of waiting, the civil servant proudly handed me a printout of the new Excel spreadsheet. However, I was in for a disappointment: the data sheet had one column with the name of each politician in the state assembly and then a column for spending—with 100% listed in each row. All politicians had spent all of their development funds. There was no variation in the spending patterns and there were no records of how they had spent it.

The only information kept in the secretariat was what percentage—100%—of the allocated funds had been spent. I thanked the civil servant for his help, and left the secretariat feeling miserable.

CONCLUSIONS

As the above examples show, data gathering often requires much of the same use of persistence, patience, local language skills, and relationship building as other forms of fieldwork. A main lesson from my work has been that it can be a huge advantage to work together with others in the field. Traveling with others can make the work safer, easier, and more enjoyable. I also learned never to rely on getting data from one source, but to try various avenues. This is important for ensuring the reliability of data and for getting anything. Finally, I learned that it is essential to be polite, respectful, and to take the time to talk properly with people working with the data you are collecting.

In the previous text I have focused on some of my failures in data collection, to show that data collection could be hard work, requiring many of the same skills as other types of fieldwork. But there have also been many success stories. In many cases I obtained access to large data sources very easily. While conducting qualitative interviews, or simply spending time in the field, I also got to hear about datasets or sources of data of which I had been unaware. Overall I hope these examples, both negative and positive, serve as reminders of the importance of fieldwork for quantitative data collection. ■

REFERENCES

- Berg, Bruce Lawrence. 2003. *Qualitative Research Methods for the Social Sciences*, 5th edition. Boston: Pearson.
- Brooke Harrington. 2003. "The Social Psychology of Access in Ethnographic Research." *Journal of Contemporary Ethnography* 32(5): 592–25.
- Glesne, Corrine, and Alan Peshkin. 1992. *Becoming Qualitative Researchers: An Introduction*. White Plains, NY: Longman.
- Johnson, John M. 1975. *Doing Field Research*. New York: Free Press.
- Kellstedt, Paul, and Guy Whitten. 2013. *The Fundamentals of Political Science Research*. New York: Cambridge University Press.
- Marcus, George E. 1997. "The Uses of Complicity in the Changing Mise-en-Scène of Anthropological Fieldwork." *Representations* 59: 85–108.