

CHAPTER OVERVIEW

Chapter 7 will explore the methodology and implications of netnographic data collection in general, with a specific focus on investigative operations and sampling issues. The chapter will begin by explaining what the terms 'data' and 'data collection' mean in relation to the practice of netnography. It will then explore the implications of observer choice on the creation of data from social media traces, interactions, interviews, and researcher experiences. Then, the three main data-collection operations will be overviewed. Investigative, interactive, and immersive data-collection operations can be used together, in combination, or separately in qualitative social media research. Which operation(s) is/are used must be determined in deference to the research question and other elements of the research context. The chapter will continue to develop the ideas from Chapter 6 and will explore ethics in investigative data operations, including the right to use public versus private sites and data in research, and the need for and forms of researcher disclosure. Some of the practical elements regarding sampling and dealing with data in which message posters are anonymous or pseudonymous will then be discussed. The final subsection will foreshadow the way in which the netnographic procedures of the next three chapters will build on the base of a data-oriented methodology developed in this chapter.

UNDERSTANDING DATA

What Exactly are Data?

Data collection is a term closely related to the operations of networked computing and, as such, the meaning of the term 'data' is not obviously related to a qualitative research procedure and product. So it makes sense to begin this chapter by explaining what the terms 'data' and 'data collection' mean in relation to the practice of netnography.

Data are trustworthy pieces of information, things known or assumed to be factual, evidential, and reliable. In research, data are the informational raw materials that are selectively either co-created (e.g., through a procedure such as an experiment), observed (e.g., by an animal behaviorist observing interspecies bird sociality), or collected (e.g., by a chemist collecting samples from different points in a city's water system). The term 'selectively' is important because the act of collecting data happens as particular pieces of information are chosen to be part of a project, and then processed in the act of interpretation during the attempt to build knowledge and uncover scientific truths. As Quinton and Reynolds (2018: xvii) assert, data is central to every element of research in the digital age:

working with digital data necessitates reflection on the ethics surrounding ownership, sharing and use of digital data, which is connected to the expectations of the researcher, the participant and the reader, which is in turn connected to the expertise required of the researcher to successfully navigate those data.

Netnographic Observer Effecting

In netnography, there will be no data before a scientific observer effects their research project by classifying some form of information – usually a set of social media postings, a comment, some messages, a conversation or a thread – as data. Before classifying them as part of the project, there are only those posts and comments, messages and characters, photographs, hashtags, pinboards, pages, and online interactions. But also there are other types of information. There are the researchers' own experiences and recordings of experiences, as well as researchers' expansive observations of others' social media-related experiences with notes and reflections. And also very often there is data intentionally collected from interaction and interviews. All of these forms of information are to be considered nothing but information until they are collected. They are not yet data.

QUANTUM PHYSICS AND NETNOGRAPHY

Netnographers need to deal with the choice among ever-present options for data collection, and the way that their actions and choices shape the dataset they will use to answer their research question. The field of quantum physics has had a long philosophical connection to similar questions, calling such measurement options and their effect upon physical data collection 'observer effects'.

For this exercise, do some reading about the famous Double Slit Experiment in physics. What are the implications of this experiment for how we might understand netnography and the choices we make when doing it? Is it fair to compare conclusions from physics with those of qualitative social media research? What might we learn from such strange new worlds that can help us become better observers of social belief and behavior?

Autogenesis Prohibited

It is important to realize that data do not create themselves in some mythical act of autogenesis. Regarding facticity, Sherry said in Sherry and Kozinets (2001: 166), that 'the theory-ladenness of facts is a qualitative preoccupation'. We might similarly consider that the method-ladenness of data is a netnographic preoccupation. Information does not exist in a pristine form 'out there' as data any more than facts do, and by collecting them, we turn traces of information into data and then into facts. Facts must be gathered and measured, whether by man or machine. Facts must be observed. Data must be sampled and its rules uncovered. It has to be tested on the path to discovering something meaningful, but the complex processes of discovery that bring particular data into being are important events through which those disclosures must be interpreted.

What is data selection in science? Whether buried in rock created hundreds of millions of years ago in a geological event that is then sampled by a geologist, or locked in the genetic code of an altered fruit fly whose progeny are sampled by a geneticist, information of various types must be chosen as part of a research study to become data. As a result of their choosing, data will bear the mark of the methods, choices, and hands that chose them.

- Experimental data bears the marks of the experiment and the experimenter.
- Survey data reflects the quality of the survey questions, the sampling procedure, and the actual sample obtained.
- The sketches of birds in the wild drawn by a biologist reflect the eyes and artistic skills of their maker.

- The fieldnotes of an ethnographer studying an Amazonian tribe reflect the perspective, interests, and education of that researcher.

Every method colors the information it collects and co-creates in the act of turning it into data. Data in netnography are produced by three general types of procedural operations – investigative, interactive, and immersive – and each of them tints data into particular hues. We will briefly discuss each of these sets of procedural operations in turn.

INTRODUCING THE THREE DATA-COLLECTION OPERATIONS

Investigative Data Operations: An Introduction

Investigative operations are selective, they chose from among the vast and ever-increasing wealth of informational traces created in the act of communications between people on social media platforms, and saved in archives and real-time recordings of social media interaction. Investigative data are not directly created by the researcher's questions or writing but, instead, are created by generally unknown others and selected for various reasons by the netnographic researcher to include in the project. Often, to enable and facilitate different types of analysis, the amounts of information collected can be relatively large, and various elements of the context of their social media surroundings will be reduced or removed. Although they may have originally been written for reasons other than as a response to the particular research study, the traces that are chosen to become investigative data are shaped by the netnographer's decisions, interests, perspectives, and observer effects, and can never be completely free of them.

Interactive Data Operations: An Introduction

The second type of data operation is interactive. Distinct from investigative operations, interactive operations are the result of some sort of researcher interference, questioning, or prodding using social media or some other online or offline interrogatory tool. Interactive operations include but are certainly not limited to online interviewing, posting a public question, commenting with a query, asking via email or direct messaging, conducting mobile ethnography, using a digital diary, or utilizing a research webpage. The informational traces collected as a result of these interrogatory operations are interactive data, and they carry all the traces of the question and the questioner that we might expect of other question-based methods, such as surveys and other interviews.

Immersive Data Operations: An Introduction

Immersive operations bear the mark of their creator in an even more overt manner. Like the other operations, immersive operations evaluate and filter what is to become netnographic data from among the vast amount of social media information flowing through public and private online communications. However, immersive operations manifest a more selective set of choices. The act of collecting information for immersion tends to be much more focused on rich or 'deep' data, extensively descriptive or informative data. Immersive operations attempt to preserve as much of the original context of that data as possible. This is done through recording detailed descriptions and explanations in the netnographer's immersion journal (the netnographic equivalent of the ethnographer's fieldnotes). The data contained in the immersion journal also reflects the process of doing the research, including the various social media sites visited, clues followed, paths taken and not taken, ideas explored, original thoughts, and so on. Immersive operations are reflective and contextual, capturing the experience of doing the research, as well as attempting to preserve and curate some of its richest examples.

Qualitative Social Media Data-collection Guidelines and Examples

In general, these three different forms of operation are used either separately or together in netnography, depending upon the research question and the research conditions considered. Often, they will overlap in content and in time. For example, the immersion journal should reflect upon and explain the major data-collection choices, providing some sense of context and background missing from the investigative dataset itself. Similarly, the immersion journal would contain reflections on interviews conducted, perhaps containing personal comments and asides regarding the process of the interview and explaining the unexpected turns it may have taken. Other times, a few choice statistical analyses will be brought in as the data is calibrated. Samples are surveyed, words are processed, networks are mapped, and interesting characteristics calculated out and reported. At other times, these operations can be quite distinct.

There have been a large number of so-called 'observational netnographies' (in future, let's just call them netnographies), in which investigative data operations are the sole type of data collection utilized in the research. I have always been a big proponent of unobtrusive online observation, and often use it myself (almost always in the contexts of wider investigations that include interviews, and immersive journaling, covered in Chapters 9 and 10, respectively). One example of the 'observational' type of netnography that focuses primarily on investigative data-collection operations is Schuman et al.'s (2018: 357–8) article that seeks to understand the health issues of combat-related trauma and, in particular, to explore 'how socially disconnected

Iraq and Afghanistan combat veterans communicate narratives of their combat-related experiences' using social media. Because the purpose of the research is to understand the trauma-related themes that emerge across the military video blogs of those combat veterans, the content of these video blogs is entirely sufficient to address the guiding question. The researcher could have asked veterans why they created the videos, or why they used a particular video method rather than another, but this would only be required to answer a different question – one about each veteran's intention or motivation. Because Schuman et al. (2018) were interested in the end result, the communicative narratives and their trauma-related themes, the videos themselves provide all the data that is required. No evidence of immersive or interactive data operations is present in these observational netnographies, and as this example shows, none may be required. However, as explained in Chapter 10, even with an ostensibly observational netnography, immersive journaling practices offer a useful procedure for saving particularly choice specimens of data, reflecting upon them, and conducting the early analysis and interpretation that will assist you as you move through your data-collection process towards research communication.

Other netnographies are far more focused on interactive or immersive data. For example, Diniz et al. (2018) wanted to study how Brazilian patients with gestational trophoblastic disease (nonviable pregnancies which are highly curable, but worrisome to patients because of the potential for malignancy) used a dedicated Facebook page headed by a physician who specialized in the disease. For this study, the researchers combined observation of the online site with an interactive and survey-based research page method. The goal of the research was to explore the way that a social media site affected perception of doctor–patient relationship and emotional well-being and to understand it in relation to patients'/social media network site users' socioeconomic profile.

More reflective netnographies, such as the autonetnography of Muslim female virtual tourist experiences conducted by Rokhshad Tavakoli (2016), rely mainly on immersive data operations, and provide lengthy and meticulous autobiographical and introspective detail with little emphasis on investigative operations. Again, these are completely appropriate choices given the applicable research questions and data site contexts. Diniz et al. (2018) wanted to gain an understanding of patients' sense of their relationships with doctors, and inner phenomenological states, as well as how patients at different ages, from different regions, and with different education levels might respond differently. To accomplish these objectives, interaction was required, and to achieve it ethically, informed consent procedures were necessary. Similarly, in order to understand the intersection of female identity, religious (observant Muslim) identity, and virtual tourist identity, plunging into Second Life and recording that dive through immersive operations was an appropriate methodological choice for Tavakoli (2016). Her autonetnography of the touristic virtual journeys she personally experienced was centered upon an immersion journal-type of data collection, and did not require further investigative or interactive sorts of operations.

THE CENTRALITY OF THE RESEARCH QUESTION

As these examples attest, it makes good sense to let the type of research question you are considering dictate the relative predominance of the data operation you will use. Generally speaking, most published netnographies use investigative operations to select and filter social media data. A good number of these publications will also add interactive operations of various sorts to product data through interrogatory principles. In the past, and likely because of confusion over the role of 'fieldnotes' (a term I recommend replacing in this edition of the book), only small numbers of scholars have communicated their reflective and curatorial processes. However, immersive journaling may be good research practice to keep your data collection organized and to continuously reflect upon, analyze, and interpret particularly insightful fragments of data.

The exact approach you use for your netnography – i.e., which of the 25 different research practices you will use, in what order, and how you will combine them – should of course fit the research question asked. However, the immersion journal is there to assist you with the entire process, including your investigative and interactive data-collection operations. If there is a good reason not to disrupt an ongoing community, or if it is already providing a rich data source that can answer the stated research question – as in the example of Schuman et al. (2018) above – then investigative operations may be perfectly sufficient. If the research objective requires more phenomenological or reflective accounts, then relying solely upon interactive or immersive operations might do well enough. However, to benefit from the full range of data-collection possibilities of an ethnographic toolkit, unless there are good reasons to exclude particular kinds of data, I recommend utilizing at least two (immersive + investigative or immersive + interactive) if not all three types of data operations in your netnography.

DATA ETHICS

Ethics in Investigative Data Operations

Because investigation is the most unobtrusive and surreptitious means of collecting data in netnography, performing it ethically requires a range of understandings and informed decisions about social media user rights, data rights, reasonable compliance with platform policies, permissions from moderators, sensitive topics, vulnerable populations, data security, anonymization, and risk mitigation. These ethical elements are woven into data collection and communication operations throughout this book. In this chapter, we will begin to engage with these topics more specifically, beginning with the right to use public versus private sites and data in research, and the need for

and forms of researcher disclosure. The remainder of the topics from Chapter 6's ethical online ethnographic research flowchart (in Figure 6.3, a visualization to which this text will often refer) will be peppered throughout the remaining chapters of the book.

Public versus Private Sites and Data

Whether the site is public or private will influence the ethical procedures we should follow. Investigative data can be drawn either from a public site or a private one. Public sites should be considered to have public data, and private sites offer private data. The difference between them is as follows. In general, a private site requires registration and a log on with a password in order to access information. A public site, in contrast, is open to any browser, and does not require registration and a log on with a password. Most blogs and forums are public sites. The publicly available areas of YouTube and Reddit are public sites. The content of public sites is usually indexed by and accessible using common search engines such as Google, Yahoo!, and DuckDuckGo. Private sites might be searchable using their own proprietary search engines, but would normally require you to be logged on as a member first, before you could use their search engine.

However, as with many things in social media, there are shades of gray. For example, although it might seem that I need to be a member to search for Twitter, LinkedIn or Facebook posts, this is not so. If I want to search the public posts of Twitter, for instance, I can do so without joining the site or logging on simply by using <https://twitter.com/search-home>. In fact, if someone has blocked me from seeing their public Twitter feed, I can view it simply by logging off the public search and using the Twitter search function instead. Similarly, public profiles and posts on Facebook or LinkedIn can be searched with a number of different search engines. Therefore, public information posted on these sites and searchable using general search engines should be considered public information and these areas of the specific sites themselves can be considered public sites.

As Moreno et al. (2013: 710) have explained, United States court law has already 'examined Facebook's privacy policy and determined that individuals do not have a reasonable expectation of privacy in information they post on their Facebook pages'. The same is very likely true for their posts on any social media site. Individuals have a reasonable expectation of privacy within their homes. So, for example, someone peeking in your windows in the evening is a clear breach of that privacy. However, when an individual knowingly exposes information to the public, by posting a photo, video, or comment, there is no reasonable expectation of privacy.

Just because a site is public, however, does not mean that using data from it is ethically unproblematic. Although they may be posting publicly, the users of such sites might have some reasonable expectations of privacy (even if the courts say these expectations are not legally enforceable). As we saw from the research presented in Chapter 6 supporting the consent gap, there is considerable variety in these expectations. However, what is important for you to recognize is that you will generally see

data as public when people's privacy settings are set that way. Privacy settings being set to public has been treated as a type of proxy for the granting of consent for others to see and for researchers to use their data, presumably for good causes.

The Right to Use Public Data in Research

The second data rights consideration is concerned about whether researchers have the rights to use this mostly-public data. Is the use of social media data similar to the 'fair use' or 'fair dealing' of copyright materials in the USA and UK, respectively? That is, is it subject to certain restrictions, but otherwise effectively waived as far as research purposes are concerned (Walther, 2002)? Some of these fair use exemptions are not in effect in international law, making matters a bit more complicated. However, the European Union's GDPR (short for 'General Data Protection Regulation') legislation, which is the strictest and most elaborate general set of rules currently governing data creators' rights, is likely to be (at least for a time) the new default standard governing global research ethics.

Under GDPR, which is a complex set of regulations, research occupies a privileged position. For example, restrictions on data use and even on important ethical categories such as sensitive topics may be overridden by researchers and research organizations that implement appropriate safeguards. We will discuss those safeguards in a later section, but the key point here is that the GDPR permits the collection of public data for research purposes by legitimate research actors who are processing data for purposes that are in the public interest, including scientific and historical research purposes, as long as the researcher (or 'data controller') acts in keeping with the recognized standards for scientific research. In all, these legal guidelines strongly suggest that public data can be used by social media researchers, as long as they take appropriate safeguards. Those specific safeguards, including the most up-to-date GDPR safeguards, are baked into the process for netnography research ethics depicted in Figure 6.3 and discussed in these sections.

The Right to Use Private Data in Research

Private data is found on private sites which require registration and a log-on with a password in order to access them. Private data is usually not indexed by and accessible using common search engines such as Google, but many of the sites have their own search functions or search engines that can be accessed once a user is logged on and operating behind its firewall. Private sites are generally restricted. However, they are not completely off limits to netnographic research.

There are a number of conditions under which data can be collected from private sites and used in netnographic research. There are two important steps in this. First, the researcher must ensure reasonable compliance with the terms related by

the platform, as well as the specific site where the social interactions are occurring (e.g., a particular topic-dedicated group within Facebook or LinkedIn). Second, the researcher must gain written permission from appropriate bodies that govern the social media site.

Reasonable compliance will need to be negotiated with the platform and with relevant moderators. Numerous published netnographies talk about their research on closed groups and the efforts they made to ensure reasonable compliance. A good example is Kantanen and Manninen (2016) whose procedure took a full year and half (this is atypically long). If it is a small website or message board, such as a corporate or organizational site, compliance would likely include gaining the permission of the platform owners. Reasonable compliance may mean that the researcher must post a notice to participants on the group about the research and its purpose, how the data will be used, how privacy will be maintained, and where the final research will appear. Compliance might also mean disclosure of the researcher and their affiliations. It might mean posting a listing of verbatim (i.e., quotations drawn from social media data) that you intend to publish in the research, and allowing the people who are quoted or cited in the research to opt-out if they do not wish their posts to be used in this manner. It may not be realistic to expect Facebook or LinkedIn to grant you explicit written permission – in fact, this is highly unlikely. That does make your request to do research illegal or wrong. Your educated and informed expectations for what is reasonable are important, and you may need to justify these to further regulatory bodies. To do so intelligently, researchers are advised not only to consider these guidelines, but also to check on the relevant laws, regulations, and institutions pertaining to their countries and governing institutions.

In the second and interrelated step, the researcher must explicitly ask for and gain permission in writing from any moderators or relevant administrators of the group to gather data. There may be several moderators. If possible, you should try to determine which one, or ones, are in charge. If there is any doubt, contact all of them. Note that, in these ideal cases, the data is purely investigative; the researchers are not interacting in any way with online participants, and disclosure to the general group of social media users may be optional. Once interaction is going to occur, as it often does in netnography, the researchers also have to invoke the data ethics rules that pertain to the use of interactive data.

Researcher Disclosure

When a netnographer reaches out, projecting their voice across the digital chasm in search of other human voices, and another voice answers, then complete honesty and disclosure must prevail. It has already been said, but I want to emphasize it again, in the present: there is no prevarication in reasoning why a netnographer would want to deceive online. If you are doing research, you must disclose that activity, along with your real identity (i.e., the one on your birth certificate or driver's license).

In a comprehensive and current evaluation of the use of netnography in hospitality and tourism research, Whalen (2018: 3427) found that only 11% of published netnographic studies explicitly mentioned disclosure and/or consent, while 35% explicitly mentioned that they did not disclose their presence and 54% did not mention the topic at all.

In those cases of human contact, whether with an identified online site of social interaction, with individuals across many sites bound by similar topics, or with particular profiles of particular individuals, the netnographer should always fully disclose their presence, affiliations and intentions during any research interactions, which is especially important in the first few interactions as they established the social arrangement. During reflexive fieldnote writing and subsequent construction of the netnographic representation, the researcher has a responsibility to record those instances of disclosure and, if possible, follow them with questions of consent for the person who is the subject of the interaction. Honest researcher disclosure, without hesitation, obfuscation or deception, is the edifice upon which the research relationships in netnography are built.

If you talk to people as a part of your research, you should never, under any circumstances, engage in identity deception. Every member of your team must use their real name and disclose their actual affiliations and purpose. Tolerate in your team and yourself absolutely no deception about why you are interacting with someone or what you are doing online if you say that you are doing netnography. Even if the practice of identity play, gender mixing, and other types of altered representation is common on the site, researchers are bound by codes of research ethics that necessitate behaviors different from the norm and are compelled to disclose themselves accurately.

Disclosure and Your Profile

In the flowchart of the netnographic research ethic process shown in Figure 6.3, if you are using YouTube, Facebook, Twitter, Snapchat, WeChat, Instagram, Pinterest, Reddit, LinkedIn or some other common social media platform in order to do at least some of your netnographic fieldwork – and if your netnographic fieldwork is going to involve you communicating with other people – then I highly recommend that the fact that you are conducting research while interacting on those sites should appear prominently in your personal or user profile, and probably should also appear at regular intervals in your status updates.

In forums or more traditional online sites, the same advice about user profiles applies. In emails, your researcher status might appear in your signature line. Another possibility would be to visually communicate your different status by wearing a t-shirt or a large button, or just digitally manufacture what is needed, and have that shout out your researcher credentials in your profile picture.

Think of ethical challenges in netnography as interesting mental experiments that reveal intriguing new things about our contemporary social media consumer-entrepreneur culture. How would you gain permission to study the profiles of dead people on Facebook or LinkedIn? How would you ethically study a community of people who discuss how they relate to dead people's social media profiles? In ethnography, it has always been a huge challenge to keep the research dance separate from the rest of the researcher's life. Ethnographers regularly marry members of the culture they study. They are accepted in many different cultures because in the embodied physical world we often encounter people who are kind, accepting, and compassionate – this seems to be one of the bases of our collective humanity. In netnography, we often see a darker and more funhouse mirror-exaggerated reflection. In a way, all communications are artifacts – real or potential. When we send messages to each other, we are dancing with ghosts. In netnography, we must deal with these facts by realizing that we are also ghosts in this space, and that others may be dancing with us.

Disclosure and Cultural Entrée

If you are going to be engaging in an online conversation, this enlisting is completely analogous to joining a physical culture or conversation. Disclosure, simplicity, and clear communication are key: 'Hi, I am a researcher at XYZ University in Cityville, and I'm interested in how Moms connect on Facebook. Can I interview you about this topic on Skype or FaceTime?' The description of research can and should be a satisfying handshake. It can be a type of ethnographic entrée and ethics combined.

It is also highly desirable for the netnographer to offer some more detailed explanation about themselves while they are introducing people to the research study, beyond simply that they are a researcher conducting a study. Providing this information is relatively simple in social networking sites such as Facebook and LinkedIn. A single post can offer plenty of detail. That link can easily be sent in an email or direct message in an app. A well-written bio will bring you and your interests to life, making you a more three-dimensional figure, rather than a cartoon image of a lab-coated online researcher holding a tablet clipboard.

So, from the beginning of the research through to its end, good research ethics dictate that the netnographic researcher: (1) openly and accurately identifies themselves when contacting and communicating directly with actual human beings, avoiding all deception; (2) openly and accurately describes their research purpose for interacting with people online; and (3) provides an accessible, relevant and accurate description of their research focus and interests, preferably with some links to some official-looking webpages somewhere which are entirely legitimate, and with more links that allow people to read your bio and learn more about you and your interests.