

1

The Science of Secondary Data

Imagine

Your friend, Paul, has been racking his brain all month. He is a graduate student and wants to complete his master's thesis exploring the effects of racial injustice on the likelihood that people of color will seek mental health treatment. Though his original plan was to collect his own data, his professor gave him access to a recent data set collected by the university counseling center on mental health diagnosis, assessment, and treatment. This data source includes 3,000 students, and 1,000 of them are students of color. "But wait," Paul says to you one Wednesday afternoon while the two of you are writing at Starbucks, "I didn't think I could use someone else's data for my master's thesis? Is that even allowed?" You smile as you sip your caramel macchiato. "Sure, it's allowed. But have you ever used someone else's data before?" Paul slowly shakes his head and looks down at his laptop. "Well, then.... Are you done with classes for the day?" You ask as you lean back in your chair. Paul gives you a disturbing look as he slowly responds, "Yes... why?" "Because," you smile, "I'm about to tell you everything you need to know about using someone else's data." This chapter will introduce you to secondary quantitative and qualitative data and give you valuable examples to share with Paul.

Learning Objectives

This chapter provides a brief overview of secondary data, describes three scientific advances in using secondary data, and discusses the use of big data for knowledge acquisition. By the end of this chapter, you will be able to:

1. Discuss the science of secondary data,
2. Describe ways to be epistemologically sound in your use of secondary data,
3. Describe types of secondary data and relevant examples of secondary data, and
4. Name three scientific developments in secondary data and how they have shaped knowledge acquisition.

The Science of Secondary Data

What exactly are secondary data? What role do they play in the advancement of science? Understanding the science of secondary data begins with understanding the distinct differences between primary data and secondary data (Corti, 2012; Johnston, 2014; Kitchin, 2014). **Primary data** are new qualitative or quantitative data collected to address a fundamental research question. Primary data are also called “new” or “original” data. On the other hand, *existing data*, or **secondary data** as it is often called, are not new and undergo secondary analysis to address a research question. Secondary data are usually related to the original, primary study goals and research questions. However, they are existing data used for secondary purposes (Corti & Backhouse, 2005; Corti, Van den Eynden, Bishop, & Woollard, 2014; Trinh, 2018).

Using data collected by someone else to answer your research questions is a common and accepted practice in research and academic settings. Secondary data can range from large, publicly available data sources (e.g., U.S. Census data) to smaller, privately-owned data sources (e.g., local social service agency data for the client population). Whether they take the form of publicly available data points, papers, artifacts, or electronic documents, secondary data are valuable in health and social science research. They can provide answers to your new and preexisting research questions (Corti, 2014; Johnston, 2014).

All secondary data were primary data at some point. Adding a temporal element to the definition of primary and secondary data is critical when distinguishing between the two because you can collect primary data now that you use for secondary purposes later. There is no hard and fast rule to the length of

time that must pass before primary data are considered “secondary.” The only real distinction involves whether someone has a fundamental research question they want to answer and whether the data to address that question exists or not. For example, if you collected data two years ago and those data can help you answer a new research question today, you can use the data to answer the new research question. Perhaps the second research question might be unrelated to the original study and its intent. In this case, you would be using the existing data for a secondary purpose.

The definition of secondary data has evolved. Table 1.1 chronicles the evolution of the meaning of secondary data beginning in 1963 and ending in 2020. Under some circumstances, using secondary data to address your research questions may save you time, money, and the resources required to initiate a new primary data collection project. Though secondary qualitative and quantitative data are collected for a different purpose, certain features may help answer your research questions, especially your mixed methods questions. To lay the foundation for the remainder of the book, here I define ***mixed methods*** as the rigorous and epistemological application and integration of qualitative and quantitative research approaches to draw interpretations based on the combined strengths of both approaches to influence research, practice, and policy (Plano Clark & Ivankova, 2016; Creswell, 2015; Watkins, 2017a; Watkins & Gioia, 2015). I discuss mixed methods more in Chapter 2 of this book. So refer to this definition from time to time, as your understanding of how to incorporate secondary data into your mixed methods crystallizes, and you make stronger links between these key terms and their definitions.

Consider this scenario: You are a first-semester master’s student in public health, with hopes of doing a research project on statewide differences in alcohol consumption with your major professor. Unfortunately, your professor is swamped with a grant proposal for the next three weeks, so she has asked you to choose a research topic and locate some data to analyze. You recently overheard a classmate talking about some free and publicly available data on the Substance Abuse and Mental Health Services Administration¹ website, so you decided to check it out. Given your interests in alcohol consumption, you choose to review the Substance Abuse and Mental Health Data Archive,² where you find some online data analysis tools. After clicking “Analyze Data Online” on the left side of the screen, you are taken to the Public-use Data Analysis System (P-DAS), where you can explore basic descriptive statistics for alcohol consumption across each state in the country. You do not have much experience with statistics, so this option suits your needs and your professor’s needs. All in all, this feels like a great start to your research project on statewide alcohol consumption.

¹ <https://www.samhsa.gov/data/node/20>

² <https://datafiles.samhsa.gov/>

TABLE 1.1 • The Evolving Definition of Secondary Data

Year	Definition	Author(s)
1963	"The study of specific problems through analysis of existing data which were originally collected for another purpose."	Glaser, p. 11
1972	"Secondary data is the extraction of knowledge on topics other than those which were the focus of the original study."	Hyman, p. 1.
1976	"Secondary analysis is the re-analysis of data to answer the original research questions with better statistical techniques or answer new research questions with old data."	Glass, p. 3
1981	"A collection of data obtained by another researcher which is available for re-analysis."	Sobal, p. 149
1982	"Secondary data analysis is any further analysis of an existing dataset which presents interpretations, conclusions, or knowledge addition to, or different from, those produced in the first report on the inquiry as a whole and its main results."	Hakim, p. 1
1985	"Neither a specific regime of analytic procedures nor a statistical technique, [but] ... a set of research endeavors that use existing materials."	Kiecolt & Nathan, p. 10
1988	"Should be an empirical exercise carried out on data that has already been gathered or compiled in some way."	Dale et al., p. 3
2006	"The further analysis of an existing dataset with the aim of addressing a research question distinct from that for which the dataset was originally collected and generating novel interpretations and conclusions."	Hewson, p. 274
2007	"Even re-analysis of one's own data is secondary data analysis if it has a new purpose or is in response to a methodological critique."	Schutt, p. 4127

route

2011	"Secondary data are those data that have been made available for use by people other than the original investigators."	Pienta, O'Rourke, & Franks, p. 13
2014	"Secondary data are data made available to others to reuse and analyze that are generated by someone else."	Kitchin, p. 7
2016	"...using data from a previous or ongoing study to test new hypotheses or answer questions not initially envisioned."	Polit & Beck, p. 244
2017	"Secondary data" are data that were formerly collected for other purposes than that of the study at hand."	Prada-Ramallal, Takkouche, & Figueiras, p. 352
2018	"Secondary data analysis is commonly defined as the use of datasets, which were not collected for the purpose of the scientific hypothesis being tested."	Trinh, p. 163
2020	"Secondary data is collected by someone other than the researcher and with another purpose."	Panchenko & Samoilova, p. 1

Note: Authors often write about secondary data and secondary data analysis together (and interchangeably), noting that "secondary data" are the existing data on cases themselves. In contrast, "secondary data analyses" are the decisions and procedures associated with analyzing existing data for a secondary data.

“

... if certain features of the secondary data source could be used to advance your understanding of a research topic, why would you not consider using the secondary source?

In the previous scenario, you can access secondary data to initiate a research project about differences in alcohol consumption by state. Though you had to search for this information on a publicly available website after you overheard a classmate talking about it, you feel lucky that things worked in your favor. You have the data you

wanted, and you did not need to assemble a research team or travel state to state to collect this information. Why? This is because someone did this for you. Now ask yourself: What if you could access a qualitative or quantitative data source for a topic in your area of interest? What would you do? Would you forget about it and collect new data? Perhaps. But if certain features of the secondary data source could be used to advance your understanding of your research topic, why would you not consider using this secondary source? You can answer these questions by understanding the epistemology of secondary data.

The Epistemology of Secondary Data

Understanding the benefits of using secondary data is key to advancing your research trajectory and productivity. Knowing how to use secondary data for a **single-method study** is a first step toward integrating two secondary single-method qualitative and quantitative data sources for mixed methods purposes. In alignment with my introduction to the science of secondary data, I would also like to discuss the epistemology of secondary data. **Epistemology** is the study of the distinction between acceptable belief and opinion (Cameron, 2011; Johnson & Onwuegbuzie, 2004). It unpacks belief, truth, and justification for each. Let's consider an example of epistemology using belief. Let's say you believe that $2 + 2 = 4$, and your belief is based on truth supported by facts. If your beliefs were based on reliable information (e.g., asking someone to count the number of writing utensils on your desk and they count two pencils and two pens), it would qualify as knowledge. However, if your belief in this mathematical equation were based on unreliable opinions (e.g., asking someone to guess the number of writing utensils on your desk with their eyes closed), it would not qualify as knowledge. Data alone is not knowledge; the interpretation of data is knowledge. But how does this connect to our interest in secondary data?

How can you be epistemologically sound in your research with secondary data? How can you be epistemologically sound with secondary data in your mixed methods? First, you must be explicit about your engagement with the

secondary data (or lack thereof). If you did not help develop the original study questions or choose the eligibility criteria for the participants, then be straightforward about this and your desire to answer your research questions using data you did not help collect. In other words, be realistic about your limited knowledge of the original study development, protocol, and lack of interactions with the study participants, acknowledge that your analysis of their information is really “secondhand.”

Next, using secondary data in an epistemologically sound way means remembering that knowledge is socially and historically located within a complex cultural context. Therefore, determining the degree to which the secondary data you plan to use aligns with your anticipated research question is more of an art than a science. The focus of your alignment should be less about perfecting the fit and more about what the secondary data source adds to the larger body of work on your topic. If you cannot perfectly fit your research question onto a secondary data source, what can you discern from the secondary data source that will answer a related research question? Or what can you determine that contributes to your overall research program’s short- or long-term goals?

Finally, using secondary data in an epistemologically sound way means respecting the culture of the primary research (e.g., the culture of the research team and the culture of the research participants) and acknowledging the power dynamics between the researchers and the people being researched. The origin of the secondary data does not disappear because you are using the data for a different purpose than that of the original study team. One example is research conducted with incarcerated populations. It makes sense to acknowledge institutionalized populations when the data are recent, but sometimes, we have access to data from people in prison that are a few years old. Just because these secondary data are a few years old does not mean that the power dynamics between incarcerated individuals and the people who oversee them have changed. Be mindful of the culture and power dynamics of the researchers and those being researched, regardless of when the secondary data were collected. Doing so means you can analyze and interpret the secondary data in their social, political, and economic contexts, considering the similarities and differences between when the data were collected and the present day. Your use of secondary qualitative and quantitative data in an epistemologically sound way can strengthen your inquiry into a topic of interest and increase the utility of secondary data for single-method or mixed methods purposes.

“

Be mindful of the culture and power dynamics of the researchers and those being researched, regardless of when the data were collected.

Types of Secondary Data

Secondary data scholars have promoted its utility and value for expanding knowledge for many years. Five decades ago, Herbert Hyman (1972) affirmed that "... existing data is the extraction of knowledge on topics other than those which were the focus of the original study" (p. 1). Extending this definition beyond that of secondary data in single-methods studies, one could argue that using secondary data in mixed methods is also a worthwhile strategy for extracting knowledge. As we begin our journey toward obtaining knowledge, let us review two types of secondary data: secondary quantitative data and secondary qualitative data.

Secondary Quantitative Data

When I hear researchers discussing the possibilities of their secondary data, they usually refer to their secondary quantitative (e.g., numeric) data. **Secondary quantitative data** come in many forms, ranging from thousands of individual cases across a hemisphere to a dozen or so individual cases right there in your office building. At present, many groups collaborate to collect and archive massive amounts of data, so one may argue that it is reasonable to use those data sources to answer some of our research questions. Moreover, some scholars say that the availability of secondary data is so robust that its use is becoming more customary across various research settings (Andrews, Higgins, Andrews, & Lalor, 2012; Smith, 2008, 2011).

When the layperson hears the term secondary data, they may think about government data sources, such as the United States Census Bureau or other data collected to track population characteristics. While writing this chapter, I searched Google for "Census Data." I found it easy to access several demographic descriptors of people in the United States using the American FactFinder, Quick Facts, Data Tools, and Data Visualization. Other data sources are moving in this direction, as public and private agencies are now more transparent about their data collection, analysis, dissemination, and sharing processes.

You may be interested in locating secondary quantitative data that are thorough and provide information about the primary study. Secondary data include, but are certainly not limited to, **population-based surveys, cohort and longitudinal surveys, administrative records, and medical records**. Table 1.2 provides examples of secondary quantitative data sources you should consider in your research. You will notice there are at least four different types of secondary quantitative data sources in the table, complete with definitions and examples for each. Other secondary data sources could vary, given diverse career and occupational settings. For example, let's say you worked at a community center for seniors, and you have decided to study the types of people who use the center. The community center's multiple-choice membership surveys collected when a senior member first arrives might be an appropriate place to start. These are primary data used for community center purposes. But, for your purposes, it is secondary quantitative data. If you want to use

TABLE 1.2 • Examples of Secondary Quantitative Data Sources

Type	Definition	Example
Population-based	Population-based studies are epidemiology studies in which a defined population is followed up and observed longitudinally to assess exposure and outcome relationships.	United States Census Bureau
Cohort and other longitudinal surveys	A cohort study is a particular form of a longitudinal study (panel study) that samples a cohort (a group of people who share a defining characteristic, typically who experienced a common event in a selected period, such as birth or graduation), performing a cross-section at intervals through time. A cohort study is a panel study, but a panel study is not always a cohort study as individuals in a panel study do not always share a common characteristic.	Framingham Heart Study ¹
Administrative records	Documents related to organization functions (such as managing the facilities, finances, and personnel) and agreements, contracts, meetings, legal actions, and so on.	Staff and personnel files
Medical records	The terms medical record, health record, and medical chart are used interchangeably to describe the systematic documentation of a single patient's medical history and care across time within one healthcare provider's jurisdiction.	Patient medical records

¹You can access a complete description and details for the Framingham Heart Study at <https://www.framinghamheartstudy.org/>

these secondary data for a single-method or mixed methods project, you first need to obtain permission to access the data and use it for research purposes. Then you would need to remove the identifying information from the data (if this were not already done before you accessed the data), carefully review the demographic characteristics of the senior members of the community center, and perform a secondary analysis with those data. Naturally, depending on the size of the sample and the dependent variables you hope to examine, fascinating insight could be acquired from analyzing the surveys from the senior members of the community center.

Let's assume you were recently hired as the social scientist for a community college in your hometown. Your department chair has issued your first assignment: to analyze student demographics across the science, technology,

engineering, and math (STEM) units on campus, including reviewing students' current academic records, transcripts from their previous schools, and test scores to determine their performance levels. This is primary quantitative data for the school, but it is existing quantitative data for your secondary purposes. The first step is to obtain permission to access the data and use it for research purposes. Then you want to make sure all identifying information has been removed before you begin your analysis. Suppose your goal is to impress the department chair and justify the need for more resources for the school. In that case, you might consider analyzing the secondary quantitative data to demonstrate that math scores for STEM students at the community college have gradually increased over the past two or three years. If you were interested in justifying the need for more funding, for example, providing statistics that frame the status of the STEM students enrolled at the community college and why the funding might benefit them would be valuable. Possibilities for application are promising when you can access comprehensive secondary quantitative data. The same is true about secondary qualitative data.

Secondary Qualitative Data

Secondary qualitative (e.g., text, image, etc.) data can provide insight and help you address some of your unanswered qualitative research questions. Historically, **secondary qualitative data** have not been as popular as their quantitative counterpart. However, between the time I started writing this book and the time I finished it, three books were published on secondary qualitative analysis (Beck, 2019; Hughes & Tarrant, 2020; Largan & Morris, 2019). Some researchers have also tried to generate traction by exploring ways to maximize secondary qualitative analysis across various studies (Bishop 2007, 2009; Fielding & Fielding, 2000; Hammersley, 2009; Hinds, Vogel, & Clarke-Steffen, 1997). Given the increased respect, rigor, and credibility of qualitative data

in recent years, more scholars are interested in doing secondary analysis of existing data. Unfortunately, a deeper appreciation for secondary qualitative data lags a few years behind secondary quantitative data. This is despite the growing number of resources and

“

A deeper appreciation for secondary qualitative data lags a few years behind secondary quantitative data.

guidelines for using secondary qualitative data in research studies (see Beck, 2019; Hughes & Tarrant, 2020; Largan & Morris, 2019). Many possibilities materialize with secondary qualitative data. Table 1.3 illustrates some examples of secondary qualitative data sources you may find helpful.

Smith's (2008) book *Using Existing Data in Education and Social Research* provides a concise account of the use of secondary qualitative data in social science research. The author states secondary data "...can be numeric or non-numeric. Non-numeric, or qualitative secondary data can include data

TABLE 1.3 • Examples of Secondary Qualitative Data Sources

Type	Definition	Example
Individual interviews	One-on-one interviews between two people, one being the interviewer and the other being the interviewee	A doctor's initial appointment with a new patient
Focus groups/ group interviews	A discussion by three or more people led by a group moderator, or facilitator, on a specific topic	A group of six new mothers asked to describe their experiences giving birth at a regional hospital
Documents/ text	A written record of information	A student's school record
Audio recordings	An electronic sound file used to capture activity or human experience	A recorded telephone call made by a telemarketer to a homeowner to solicit interest in switching Internet providers
Video recordings	An electronic video file that is used to capture human experience	A YouTube video of how passers-by respond to a child in distress at the zoo
Photos and images	Pictures used to capture a concept, idea, experience, or interpretation of a human experience	Historical photographs of a town's first erected buildings, architecture, and topography plans and layout

retrieved secondhand from interviews, ethnographic accounts, documents, photographs, or conversations” (pp. 4–5). Though most secondary data tend to be quantitatively oriented, Smith reminds us of the possibilities of secondary qualitative data in our education and social research endeavors.

Let's say you are a tenure-track assistant professor conducting qualitative interviews to learn more about the experiences of parents whose children are in the foster care system. You hope to understand the biological parents' knowledge, attitudes, and beliefs about foster care. You begin by working with your ethics board to ensure that interactions with the study participants are handled appropriately, and then you collect the data. You also de-identify the data and make the appropriate ethical considerations for data-sharing, as instructed by your ethics review board.

Once the answers to your questions are finalized (and if you were willing), you might consider sharing the de-identified qualitative data with colleagues and students to help answer their related research questions. For example, let's assume I am a third-year Ph.D. student in your department. I am interested in using your data to learn about the biological parents' demographics and their beliefs about the foster care system. You might consider sharing your

primary data for my secondary purposes. As a first step toward addressing my research question, I might schedule a meeting to discuss the participants' demographic information you collected. Though this was not a primary focus of your research, participants might have provided this information in a short survey or during the qualitative interviews, so it would be great to maximize your secondary data to advance my research. Studies that use secondary data are not excluded from an ethics review, so be sure to check with the ethics review committee to learn about the process for undergoing an ethics review for a study that uses existing data for secondary purposes.

What if I conducted a primary study in which I video-recorded 32 patient-physician interactions (16 with male patients and 16 with female patients), and I used these data to understand the role that primary care physicians play in addressing their patients' mental health concerns. Some patients do not go directly to mental health professionals when dealing with mental health conditions. Instead, they may seek help from their primary care physician about the outward effects of mental health conditions manifesting as physical health conditions (e.g., headaches, body aches, insomnia, etc.). Let's assume I used these data to answer my research questions about the mental health implications of such interactions, and that was all I hoped to do with the videos. You might express interest in using my videos as a secondary qualitative data source to explore the gender dynamics in patient-physician interactions in clinical settings. Again, I should note that specific permissions are needed to access such data since videos of healthcare patients are involved. So an ethics review would be required, and you will need to outline a process for obtaining permission to use the videos for a secondary purpose. Nevertheless, based on the types of questions I included in my primary study protocol, you might be able to glean some interesting information to help you address your research questions. See Table 1.4 for other real-world examples of purpose statements, research questions, and primary and secondary data.

Scientific Developments With Secondary Data

Secondary data have practical applications for both single-method and mixed methods research. As you will note throughout this text, while I recommend scholars consider using secondary data for mixed methods studies, I recognize that it is not always obvious when a mixed methods study can benefit from secondary data. Likewise, a more persuasive rationale may be needed for how secondary data are used to acquire knowledge. Many scholars have advocated using secondary data and analyzing secondary quantitative data (Hyman, 1972; Johnston, 2014; Smith, 2008). Rather than overstating the points these earlier scholars made regarding the value of secondary data in single-method studies, here I focus on three scientific developments of secondary data and how they can help you think about expanding your research, be it single-method or mixed methods focused.

TABLE 1.4 • Real-World Examples for When to Use New Data Versus Secondary Data

Purpose Statement	Research Question	Study Context	New vs. Secondary Data (Why?)
This study aims to test the Health Belief Model by comparing male and female smokers who are in our smoking cessation program.	Are there sex differences in smoking cessation as demonstrated by the program participants?	Quantitatively oriented; unique to a smoking cessation program.	New data (because this inquiry is unique to the participants in this smoking cessation program)
The purpose of this qualitative study is to explore the counseling experiences of first-generation first-year students at the University of California, Berkeley.	What are some advantages and disadvantages to counseling for first-generation students?	Qualitatively oriented; unique to the first-generation freshmen at the University of California, Berkeley.	New data (because this inquiry is unique to the first-generation students at the University of California, Berkeley)
This study aims to examine the depressive symptoms of African Americans, Caribbean Blacks, and Whites in the United States.	Are there racial/ethnic differences in depression symptoms for adults in the U.S.A.?	Quantitatively oriented; several national data sources exist, such as the NCS ^a and the NSAL ^b .	Secondary data (because current, national data to address the research question already exist)
This phenomenological study aims to understand the religious practices of female clergy at three Pentecostal Churches in the city.	What are the unique experiences of female clergy in the Pentecostal Church?	Qualitatively oriented; your colleague is a religious scholar who has collected these data from the three churches.	Secondary data (because current, localized data to address this research question already exist)

^a The National Comorbidity Survey (NCS) was the first nationally representative survey of the prevalence and correlates of psychiatric disorders and treatment in the United States (1991 and 1992). There are NCS extensions, including a 10-year follow-up of the baseline NCS sample and a replication of the NCS (2001 and 2002).

^b The National Survey of American Life (NSAL) is the first national survey to investigate ethnic differences within the Black population by including significant numbers of Black Caribbeans.

Knowledge Pursuit as a Scientific Development of Secondary Data

The first scientific development of secondary data has allowed scholars to pursue knowledge in unique and sophisticated ways. As a health or social science scholar, not only do you have your research program but you are also encouraged to think about the ways that other scholars conduct the same or similar work. There are several benefits to using secondary data for one's research projects, such as exposure to new and different research foci, extending your understanding of a topic beyond the scope of your geographic area, and your accessibility to the breadth (or depth) provided by learning from new research participants.

But one of the most fundamental benefits of secondary data is its ability to challenge you by extending your pursuit of knowledge above and beyond your current methodological boundaries and methods margins. This is your greatest

“

This is your greatest gift from the original research investigators: an opportunity to think about a research question outside of your usual conceptual and practical parameters.

gift from the original research investigators: an opportunity to think critically about a research question outside your usual conceptual and practical parameters. Furthermore, countless learning opportunities come along with using secondary data. The ability to work with data that you did not collect allows you to gain an insider's view of a complete

data set, its strengths and weaknesses, and its ability to address a series of research questions associated with your research interests. This insider's view of a complete project can be a formative experience for new researchers who aspire to lead their own primary data collection projects someday.

Technology as a Scientific Development of Secondary Data

The second scientific development in secondary data is the growth and expansion of technology. Advances in technology have allowed us to streamline our data management during three vital research stages: data collection, data storage, and data-sharing. First, technological advances in data collection have far exceeded those of the researchers who came before us. While early researchers had to depend on face-to-face interactions and the telephone to collect their data, we now have the Internet as a means for collecting both qualitative and quantitative data for our research purposes. Web- and email-based surveys are frequently used by scholars in the health and social sciences and have resulted in more streamlined data collection methods for studies that require quantitative (e.g., online survey) or qualitative (e.g., video-recorded interview) data.

Second, with access to space on the Internet and on the “cloud,” unlimited data storage is a possibility, not only for the Ph.D.-trained scientist who has 1,000 data files to store but also for the stay-at-home parent who has 1,000 high-quality photos of their children to store. Simply put, because data storage options nowadays are maximized or unlimited, “cloud-based” storage has become a way of life for us all, including scholars working with qualitative and quantitative data for research purposes. Finally, data-sharing is much easier now compared to 50 years ago. More people than ever before are storing and managing their data sources on the Internet and devising data-sharing plans so that others have access to their data files. Sophisticated data-sharing plans have become a requirement for federally funded research projects; therefore, there is an expectation to access qualitative and quantitative data collected using federal funds. For example, the National Institutes of Health now require applicants to include data-sharing plans in their grant proposals (<https://grants.nih.gov/policy/sharing.htm>). Aligned with this, health and social sciences research topics are becoming more complex, and now, these complexities are being shared using technology and data-sharing mediums. An example of this is how some Internet-based scientific journals allow authors to publish supplemental materials to accompany their accepted publications.

“

While early researchers had to depend on face-to-face interactions and the telephone to collect their data, we now have the Internet as a means for collecting both qualitative and quantitative data for our research purposes.

Big Data as a Scientific Development of Secondary Data

The last of the three scientific developments in secondary data is big data. As its popularity evolves, **big data** are becoming an essential component of research to improve population health and well-being. But you may be asking yourself, what exactly are big data? Big data are “...data that exceed the processing capacity of conventional database systems” (Lauzon, 2012, p. 4). In other words, “big data” are too big, move too fast, or do not fit the traditional database storage and management systems structures. To gain value from big data, some scholars suggest we choose an alternative way to process it, understand it, and apply it to our world (Dumbill, 2013; Mayer-Schonberger & Cukier, 2013); the same can be said about considering big data as a scientific development in secondary data.

In the current scientific discourse, big data are used by different industries, disciplines, and platforms (including business, healthcare, social science, and information technology). The current literature is broad and deep, and while big data are a topic of growing interest, it lacks an agreed-upon definition. For this text, I acknowledge two things: (1) that big data are collected at incredible rates, and (2) that big data users are finding more and more ways to capitalize on their

bigger, faster, and more substantial ambitions for data acquisition, management, and use. Some examples of big data are data found in mobile devices, the Internet, social networking sites, political polling, and entertainment (Brown, Chui, & Manyika, 2011; Heafner, Fitchett, & Knowles, 2016).

Big data and mixed methods that use secondary data are similar in their introduction to the research world, acceptance, and application. Because the core contribution of big data is how it can help us perform at a larger scale, it can also make a variety of contributions to how we understand and apply secondary data to mixed methods to answer research questions. Just as big data researchers are seeking to extract more abundant insights and create new areas of foci as they delve deeper into phenomena of interest, so too are mixed methods researchers who are attempting to reap the benefits of secondary qualitative and quantitative data and integrating them so that the strengths outweigh the weaknesses. This will allow us to develop a complete picture of a problem under investigation.

“

[Big data's] appeal is not necessarily its quantity, but rather how our understanding of the world is shaped and becomes a direct by-product of how we analyze and interpret those data.

The appeal of big data is not necessarily its quantity, but rather how our understanding of the world is shaped and becomes a direct by-product of how we analyze and interpret those data. Simply put, to understand big data is to acknowledge the importance of predictions in our world; big data is not about teaching

a computer to think like a person but about applying sophisticated math equations to massive quantities of data so that we can infer probabilities that improve the way we function as a society (Mayer-Schonberger & Cukier, 2013). Thus, big data's contributions to mixed methods have only just begun. We have not yet scratched the surface of advancing research methods with big data and how they can radically change how we think about secondary data in mixed methods.

Summary

This chapter introduced the science of secondary data, described ways to be epistemologically sound when using secondary data, described secondary data types and relevant examples of secondary data, and described three scientific developments in secondary data and how they have shaped knowledge acquisition. The science of secondary data must begin with understanding the distinct differences between primary and secondary data. While primary data are qualitative or quantitative data collected to address a primary research question, secondary data are used for a secondary purpose: to address a new, sometimes related, research question. Now that you have read this chapter, you should be able to describe ways to be

epistemologically sound in your use of secondary data. There are at least three ways to conduct epistemologically sound research with secondary data. First, be realistic about your limited knowledge of the original study and acknowledge that your analysis of their information is really “second-hand.” Second, remember that knowledge is socially and historically located within a complex cultural context. Third, respect the culture of the primary research and acknowledge the power dynamics between the researchers and the people being researched. Secondary data are used to address a secondary research or evaluation question that may or may not be related to the original research goals and questions. Finally, you should be able to name three scientific developments in secondary data and how they have shaped the acquisition of knowledge: the pursuit of knowledge, technology, and the use of big data. Each has contributed in its unique ways, but all have advanced how we use secondary data in single-method and mixed methods research.

Chapter 1 Application Questions

1. What are secondary data, and what are some examples of secondary data in the health and social sciences?
2. What are some ways to be epistemologically sound in your use of secondary data?
3. What are three scientific developments in secondary data, and how have they shaped knowledge acquisition?
4. What are big data and its contributions to mixed methods with secondary data?
5. What are some real-world examples that may require the use of secondary data?
6. How does the pursuit of knowledge influence your decision to use secondary data in your single-method or mixed methods research?
7. Can technology play a role in your research that uses secondary data? If so, how can data collection, storage, and sharing be maximized in your project using technology?
8. What role, if any, does big data have in your research? Can you see big data making an essential contribution to your study as secondary data for either a current or future single-method or mixed methods project? Why or why not?